PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
MESTRADO EM CIÊNCIA DA INFORMAÇÃO
UNIVERSIDADE FEDERAL FLUMINENSE

SUEMI HIGUCHI

Representação do conhecimento e modelagem conceitual de ontologia no domínio da História do Brasil Contemporâneo

Niterói

2012

SUEMI HIGUCHI

Representação do conhecimento e modelagem conceitual de ontologia no domínio da História do Brasil Contemporâneo

Dissertação de Mestrado apresentada como requisito para obtenção do título de Mestre em Ciência da Informação no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal Fluminense.

Orientador: Prof. Dr. Carlos Henrique Marcondes

Co-orientador: Prof. Dr. Renato Rocha Souza

Niterói

2012

Ficha Catalográfica elaborada pela Biblioteca da Escola de Engenharia e Instituto de Computação da UFF

H638 Higuchi, Suemi.
Representação do conhecimento e modelagem conceitual de ontologia no domínio da História do Brasil Contemporâneo. – Niterói, RJ: [s.n.], 2012.
173 f.

Dissertação (Mestrado em Ciência da Informação) – Universidade Federal Fluminense, 2012.
Orientador: Carlos Henrique Marcondes.
Co-orientador: Renato Rocha Souza.

1. Representação do Conhecimento. 2. Web Semântica. 3. Modelagem Conceitual. 4. Ontologia. 5. História do Brasil Contemporâneo. I. Título.

CDD 020.1 (21.ed)

SUEMI HIGUCHI

Representação do conhecimento e modelagem conceitual de ontologia no domínio da História do Brasil Contemporâneo

Dissertação de Mestrado apresentada como requisito para obtenção do título de Mestre em Ciência da Informação no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal Fluminense

Aprovada em: ____

Banca examinadora:

____

Prof. Dr. Carlos Henrique Marcondes (Orientador e Presidente)
Programa de Pós-Graduação em Ciência da Informação, Universidade Federal Fluminense (UFF)

____

Prof. Dr. Renato Rocha Souza (Co-orientador)
Programa de Pós-Graduação em Matemática Aplicada, Fundação Getulio Vargas (FGV)

____

Prof[a]. Livre-Docente Dra. Hagar Espanha Gomes (Membro titular)
Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

____

Prof[a]. Dra. Maria Luiza de Almeida Campos (Membro titular)
Programa de Pós-Graduação em Ciência da Informação, Universidade Federal Fluminense (UFF)

____

Prof[a]. Dra. Rosa Inês de Novaes Cordeiro (Suplente)
Programa de Pós-Graduação em Ciência da Informação, Universidade Federal Fluminense (UFF)

Niterói

2012

AGRADECIMENTOS

Considerando esta dissertação como resultado de uma caminhada que não começou no Mestrado, agradecer pode não ser tarefa fácil, nem justa. Para não correr o risco da injustiça, agradeço de antemão a todos que de alguma forma passaram pela minha vida e contribuíram para a formação de quem sou hoje. Sintam-se muito queridos.

E agradeço, particularmente, a algumas pessoas pela contribuição direta na construção deste trabalho, e outras, por terem ajudado de alguma forma, no processo:

Ao meu orientador Prof. Dr. Carlos Henrique Marcondes por seu apoio e inspiração no amadurecimento dos meus conhecimentos e conceitos que me levaram a execução e conclusão desta dissertação.

Ao Prof. Dr. Renato Rocha Souza, que, com carinho, aceitou a tarefa de me co-orientar e tão generosamente dividiu comigo não apenas sua erudição, mas também sua amizade, experiência e sabedoria de vida. Nossas reuniões de trabalho sempre me acresceram também como ser humano.
À Profª. Maria Luiza de Almeida Campos e Profª. Hagar Espanha Gomes, pelas valiosas contribuições e discussões que ajudaram a alimentar meu encanto pela Ciência da Informação.

Ao Centro de Pesquisa e Documentação de História Contemporânea do Brasil (CPDOC/FGV), instituição da qual faço parte desde 2001 e muito me orgulho, e seu diretor prof. Celso Castro, pelo apoio e compreensão nestes dois anos de curso.

Aos amigos da Turma do Mestrado 2010, sempre juntos, na angústia e no alívio. Em especial à Deborah Motta Ambinder e Caroline Brito, pela força, carinho e amizade que se construiu para além dos espaços da universidade.

Mesmo longe, Daniela Lucas foi uma amiga próxima com quem compartilhei aprendizado, conhecimento na área de Ontologias e lições de vida. Dani é uma das pessoas mais iluminadas e serenas que tive o prazer de conhecer.

Na FGV, Alexandre Rademaker ajudou a construir o meu interesse pelas questões que permeiam esta pesquisa. Dos nossos primeiros e descompromissados bate-papos, até os projetos de ponta que compartilhamos hoje, serei sempre grata pela diferença que fez para o meu conhecimento.

No CPDOC aprendi que nem só de trabalho vive o homem, principalmente quando a rotina nos obriga à convivência com pessoas estranhas e diferentes. Mas tenho sorte porque dentre estas pessoas, encontrei espécies raras que tornam meus dias leves e divertidos, e temos sempre um motivo na manga para comemorações e almoços especiais. Aos queridos Jimmy

Medeiros, Juliana Gagliardi, Vanessa Matheus, Vivian Fonseca, Aline Santiago e Cláudio Pinheiro, juntos e misturados.

Às três pessoas com quem tenho a felicidade de conviver também no trabalho, mas principalmente fora dele, e que de maneiras diversas fazem enorme diferença na minha vida profissional e afetiva: Kely Lobo Neves, Bianca Freire-Medeiros e Márcio Grijó Vilarouca. O que seria de mim sem vocês? O que seria do Márcio sem mim para praticar seu *bullying* diário? Queridos, vocês são presentes na minha vida, guardo para sempre no coração.

Há alguns anos fui privilegiada com pessoas muito queridas, onde sobram carinho e amizade: Weykman, Rogério Koscheck, Marisa, Nelsinho, Anderson e Cléber. Além, é claro, da Keké e do Edmilson. É sempre festa.

Agradeço também de forma especial aos meus queridos amigos Dani Cavaliere, Melissa, Sésiom, Tiago, Juju, Jeane, Dirô e André pelos fins de semana sempre preenchidos com muita gastronomia, risadas e Master. Nos momentos difíceis, não nos largamos nunca.

À minha grande família, pelo incentivo e carinho. Ao Franki, à Leika, ao Phill, Yuuki e Yuna, e em especial aos meus pais, Emico e Teruo Higuchi, porque o sacrifício deles para se chegar aqui não foi pequeno. Amo vocês e devo muito.

Acredito que Deus tem agido na minha vida colocando no caminho as pessoas certas nos momentos certos, como se fosse uma grande ‘coincidência’. Mas eu sei que é dessa forma que Ele cuida e me ajuda a seguir pela realização dos desejos do meu coração.

“Mas é do buscar e não achar
que nasce o que eu não conhecia”

(Clarice Lispector)

RESUMO

Uma ontologia é um sistema particular de categorias que versa sobre certa visão do mundo. No contexto da web semântica seu propósito é classificar as entidades de um domínio, definir seu vocabulário e explicitar as relações de significado a fim de que possam ser interpretadas computacionalmente através de regras lógicas. Neste trabalho são apresentados os desafios para a modelagem de uma ontologia no domínio da História e as possibilidades de associação dos registros históricos aos seus contextos de significados através dos construtos da web semântica. Em particular, analisam-se metodologias e experiências nesta seara da representação do conhecimento, endereçando um estudo de cunho exploratório sobre modelagem conceitual de uma ontologia no domínio da história do Brasil recente.

Palavras-chave: Representação do Conhecimento, Web Semântica, Modelagem Conceitual, Ontologias, História do Brasil Contemporâneo.

ABSTRACT

An ontology is a particular system of categories that deals with a certain vision of the world. In the context of semantic web its purpose is to classify the entities of a domain, define its vocabulary and explicit the relationships of meaning so that they can be computationally interpreted by logical rules. This study presents the challenges for modeling an ontology in the field of History and the possibility of associating historical records to their contexts of meaning through the semantic web constructs. In particular, it examines methodologies and experiences in this field of knowledge representation, addressing an exploratory investigation on conceptual modeling of an ontology in the domain of contemporary history of Brazil.

Keywords: Knowledge Representation, Semantic Web, Conceptual Modeling, Ontologies, Contemporary History of Brazil

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS E ABREVIATURAS

ACE - Automatic Content Extraction Evaluation

BBC PO - BBC Programmes Ontology

BIBO - Bibliographic Ontology

CC - Ciência da Computação

CI - Ciência da Informação

CIDOC CRM - CIDOC Conceptual Reference Model

CPDOC - Centro de Pesquisa e Documentação de História Contemporânea do Brasil

CRG - Classification Research Group

DHBB - Dicionário Histórico-Biográfico Brasileiro

DL - Description Logic

DOLCE - Descriptive Ontology for Linguistic and Cognitive Engineering

EAD - Encoded Archival Description

EDM - Europeana Data Model

FGV - Fundação Getulio Vargas

HIST - Ontologia de História

HTML - HiperText Markup Language

IA - Inteligência Artificial

ISBN - International Standard Book Number

ISO - International Organization for Standardization

KIF - Knowledge Interchange Format

KOS - Knowledge Organization System

LOD - Linked Open Data

LODE - Linking Open Descriptions of Events

MARC - Machine Readable Cataloging

N3 - Notation 3 RDF

OAI-ORE - Open Archives Initiative Object Reuse and Exchange

OWL - Web Ontology Language

PLN - Processamento de Linguagem Natural

RDF - Resource Description Framework

RDF-S - Resource Description Framework Schema

SEM - Simple Event Model

SKOS - Simple Knowledge Organization System

SNaP - Simple News and Press Ontologies

SUMO - The Suggested Upper Merged Ontology

SW - Semantic Web

SWS - Semantic Web Services

TED - Technology, Entertainment, Design

TEI - Text Encoding Initiative

TGN - Getty Tesaurus of Geographical Names

TGT - Teoria Geral da Terminologia

TOVE - Toronto Virtual Enterprise

UFO - Unified Foundational Ontology

UML - Unified Modeling Language

URI - Uniform Resource Identifier

VICODI - Visual Contextualization of Digital Content

W3C - World Wide Web Consortium

XML - eXtensible Markup Language

YAGO - Yet Another Great Ontology

CONVENÇÕES

- Nos trechos de código utilizou-se fonte de letra *Courier* para destacá-los do restante do texto:

```
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
```

- Da mesma forma, na apresentação de classes e propriedades de ontologia, a fonte *Courier* é adotada. Classes sempre iniciam com a letra maiúscula e propriedades com a letra minúscula:

```
Evento; temDataInicio
```

- Para referenciar classes e propriedades de ontologias externas, a sigla pela qual a ontologia é reconhecida precede a classe ou propriedade em questão:

```
edm:Event; edm:happenedAt
```

SUMÁRIO

1 Introdução

Temos testemunhado de forma progressiva a passagem dos formatos analógicos para os digitais. Caminhamos para o estado em que todo produto da atividade intelectual humana já nasce digital ou logo estará disponível em algum formato eletrônico, e o predomínio desta modalidade traz em si mudanças não só na forma como elaboramos e expressamos o conhecimento, mas também na maneira como o identificamos e selecionamos para uso.

Com o barateamento dos recursos de armazenamento e a disseminação cada vez maior da web como espaço para compartilhamento e troca de informações, os temores que Vannevar Bush identificou há mais de 65 anos[1] seguem desafiando pesquisadores, cientistas e usuários numa busca permanente por formas de fazer frente à expansão da informação e sua apropriação (BUSH, 1945). No âmbito da gestão de informação as tecnologias disponíveis sempre foram os balizadores na escolha de modelos e métodos a serem adotados em suas inúmeras aplicações. O problema é que diante do crescimento desordenado e caótico da web, as soluções de software existentes falham ao providenciar qualquer significado ou entendimento à grande massa de dados, tornando a localização da informação relevante a maior dificuldade da atual economia da Informação (MARCONDES & CAMPOS, 2008).

Em 1951, Calvin Mooers já antevia estes problemas identificando nos estudos emergentes sobre recuperação da informação as três questões fundamentais: i) como descrever intelectualmente a informação; ii) como especificar a busca; iii) que sistemas, técnicas ou máquinas devem ser empregados (MOOERS, apud SARACEVIC, 1996).

Mesmo conjugando tarefas que não seriam possíveis anteriormente, como buscas por palavras-chave, as versões digitais ainda estão longe do ideal, pois não transpõem as limitações que subsistem ao tentarmos localizar e manipular as informações de forma mais inteligente. Relações entre conceitos no interior de um texto não aparecem nestas buscas, apenas se é capaz de encontrar as instâncias desses conceitos. Por exemplo, se duas instâncias, personagem A e personagem B, são recuperadas, somente com a leitura do documento será possível determinar se estas duas pessoas se relacionam, que tipo de relacionamento possuem,

[1] Vannevar Bush, cientista do MIT e chefe do departamento de pesquisa e desenvolvimento científico americano na década de 1940, identificou na época o que ficou conhecido como problema da *explosão informacional*, ou seja, do irreprimível crescimento exponencial da informação e de seus registros, particularmente em ciência e tecnologia (SARACEVIC, 1996).

e durante que período ele se deu. Ou seja, a captura não vai além de palavras sem significado e deixa de revelar informações relevantes em contextos específicos.

A comunicação humana necessita da contextualização para que a mensagem seja recebida sem ambiguidades. Por definição, dados fora de contexto são apenas fragmentos soltos de informação, pouco úteis para a compreensão de um dado discurso. Somente quando a informação se encontra estruturada numa rede de conexões que façam sentido, é possível compreendê-la. Como observou Ian Cornelius, "cada bit de informação só é informação se a entendemos no contexto no qual está empacotada, o qual nos permite interpretá-la" (CORNELIUS 1996, apud CAPURRO, 2003).

Interpretar textos de forma automática ainda é um desafio para a comunidade científica, que vem sendo enfrentado com a ajuda de técnicas desenvolvidas principalmente no campo da Inteligência Artificial. Aplicações de aprendizagem de máquina e mineração de texto são empregadas para a localização de estruturas de semelhança e padrões em grandes conjuntos de dados. Na área de processamento de linguagem natural (PLN) o desafio está em fornecer aos computadores capacidade de "entender" textos escritos em linguagem humana para uma série de aplicações automatizadas. Isso inclui realizar a análise sintática, léxica e morfológica dos seus sintagmas, convertendo as ocorrências em representações formais.

Nesse campo fértil de pesquisa, o projeto da Web Semântica acena com um conjunto de tecnologias capazes de atribuir anotações semânticas às informações. A expressão *Machine understandable information* sintetiza a visão de Tim Berners-Lee[2], para quem a web se estenderia da apresentação pura e simples dos dados para seu compartilhamento mais significativo, informação 'compreensível' tanto pelo homem quanto pela máquina (BERNERS-LEE, 1998). Para tornar realidade, parte-se de um princípio fundamental neste processo, o da criação de modelos conceituais bem definidos capazes de representar sem ambiguidades o conhecimento desejado.

Modelos conceituais são artefatos produzidos com o objetivo de representar uma dada porção da realidade segundo uma determinada conceituação (GUIZZARDI, 2000), e nesse sentido as ontologias despontam como estruturas consistentes de representação. No plano filosófico o termo se refere ao estudo metafísico da natureza do ser e sua existência (SOWA, 1999; SMITH, 2002) ou um sistema particular de categorias considerando certa visão do

---

[2] Tim Berners-Lee: engenheiro e cientista da computação britânico, criador da World Wide Web e precursor da Web Semântica.

mundo (GUARINO, 1998). Na ciência da computação seu propósito é classificar as entidades de um domínio, definir seu vocabulário e explicitar as relações de significado existentes (SMITH, 2002) a fim de que possam ser interpretadas computacionalmente através de regras lógicas. Quanto mais relações de significado, ou seja, riqueza semântica existir no domínio concebido, mais superior será o poder de inferência. E as inferências servem para aumentar a precisão[3] da informação pesquisada.

1.1 Pressupostos e questões de pesquisa

O presente trabalho visa a endereçar estas questões acerca da representação do conhecimento no domínio da História. Como não poderia deixar de ser, a informação histórica transmitida em formato eletrônico padece de todas as limitações mencionadas anteriormente. Instituições de memória, centros de pesquisa e de documentação vêem na Internet um canal oportuno para a disseminação de seu acervo e produção, no entanto sabem que ao mesmo tempo em que o acesso é ampliado, a ausência de um serviço de referência adequado dificulta a assimilação do conteúdo por parte dos usuários remotos. São páginas e páginas de informação não estruturada, sistemas de busca meramente sintática, ausência de padronização terminológica e outras limitações que acabam por esconder, mais que revelar, o conhecimento do público.

Em particular, textos narrativos históricos contêm certos elementos que caracterizam o gênero; eles são geralmente organizados sobre uma linha do tempo e partilham de conceitos chave tais como pessoas, lugares e eventos. A questão está em perceber como estes conceitos devem ser capturados no contexto da escrita da história: que propriedades garantem a sua identidade no tempo e no espaço, quais carregam as mudanças?

A história como domínio do conhecimento possui aspectos que podem ser problemáticos do ponto de vista de modelagem de ontologia. Segundo Gadamer (apud CAPURRO, 2003), enquanto as ciências naturais são movidas por leis universais e invariáveis, as ciências humanas tentam alcançar o sentido da verdade através da interpretação dos fenômenos. Noções históricas podem ser definidas de forma vaga, não consensual e abertas a múltiplas interpretações; questões sobre *onde*, *quando*, *quem* e *o que*, por vezes não

[3] O sucesso dos Sistemas de Recuperação da Informação (SRIs) está vinculado a dois índices: revocação e precisão. O primeiro mede a capacidade que um sistema possui de recuperar documentos pertinentes, enquanto o segundo avalia, dentre os documentos recuperados, quais correspondem realmente à necessidade de informação do usuário (SOUZA, 2006).

são consideradas mais importantes do que *por que* (NAGYPAL, 2005). Ou seja, historiadores tendem a estar mais interessados nos fatos em contextos (conhecimento) do que tão somente nos simples fatos (informação).

Diante disso, questiona-se se codificar conhecimento histórico usando os construtos relativamente simples das linguagens de representação poderia seguir o mesmo raciocínio das outras ciências. Que princípios metodológicos orientariam o desenvolvimento de uma ontologia no domínio da História? Como as estruturas ontológicas representariam apropriadamente a dinamicidade temporal presente na História? Quais conceitos deveriam ser integrados à ontologia?

Este trabalho visa, sobretudo, elucidar estas questões de pesquisa.

1.2 Objetivos

1.2.1 Objetivo geral

- Investigar as bases para a modelagem de uma ontologia no domínio da história do Brasil recente, identificando as contribuições teóricas e metodológicas que melhor endereçam estas bases.

1.2.2 Objetivos específicos

- Identificar os desafios da representação do conhecimento no domínio da História, caracterizando as particularidades percebidas em relação a outros domínios, em especial das ciências ditas naturais e exatas;

- Propor um esquema inicial de ontologia no domínio da história do Brasil, privilegiando um recorte temático contemporâneo tendo por base os pressupostos teóricos e metodológicos visitados, e que dê sustentação ao objetivo geral.

1.3 Justificativa

O trabalho se justifica considerando três dimensões: a relevância para a Ciência da Informação, a relevância teórica e a relevância social.

Segundo Borko (1968), a Ciência da Informação compreende em seu corpo de conhecimento pesquisas sobre as formas de representação da informação tanto em sistemas naturais quanto artificiais, o uso de códigos para a transmissão eficiente das mensagens e o estudo dos recursos e dispositivos técnicos, como computadores e seus programas. Ela é parte ciência pura, através da pesquisa dos fundamentos, e parte ciência aplicada, ao desenvolver soluções por meio de produtos e serviços.

Outros autores (VICKERY, 1986; CAPURRO, 1991; ALVARENGA, 2003) também consideram os estudos sobre representação como primordiais às atividades de investigação do campo, relacionando-os à produção dos registros de conhecimento, à organização dos sistemas de informações documentais e ao acesso às informações pelos usuários.

Estas questões são nucleares para a Ciência da Informação desde há muito tempo. Não à toa, os relatos da literatura de língua inglesa da década de 60 e boa parte da década de 70 mostram uma

> “ciência da informação atrelada à teoria matemática da comunicação; empolgada com a nascente automação de sistemas de recuperação da informação e de bases de dados; voltada para os problemas de *semântica*, visando à *representação da informação*; preocupada com os primeiros estudos de relevância e medidas de desempenho e avaliação de SRIs[4]”. (BRAGA, 1995, p. 4).

Para alguns, a Ciência da Informação pode ser considerada, eminentemente, uma ciência da representação, onde seus profissionais desempenham atividades em grande parte voltadas para SRIs. “Em cada registro de conhecimento, em seus metadados, e em cada necessidade de informação, está implícita a necessidade de representação para que seja possível a comunicação de subjetividades” (SOUZA, 2008, p.2).

---

[4] SRIs ou Sistemas de Recuperação de Informação.

Assim, esta pesquisa vai ao encontro do que a Ciência da Informação se propõe como disciplina, evidenciando a sua importância como suporte de investigação frente às questões colocadas, buscando no ferramental tecnológico disponível a base para este processo.

> "Ciência da Informação é um campo dedicado às questões científicas e à prática profissional voltada para os problemas da efetiva comunicação do conhecimento e de seus registros entre os seres humanos, nos contextos social, institucional ou individual do uso e das necessidades de informação. No tratamento destas questões são consideradas de particular interesse as vantagens das modernas tecnologias da informação." (SARACEVIC, 1996)

O levantamento de debates e teorias advindas de outras áreas, como a Ciência da Computação, Linguística e História, além de fortalecer a posição multidisciplinar característica da Ciência da Informação, favorece o tratamento mais adequado dos registros históricos e permite melhorar a forma como identificamos e representamos essas fontes.

Além disso, iniciativas de construção de ontologias no domínio das humanidades, ainda que venham crescendo em número nos últimos tempos, são mais tímidas que nas ciências exatas e naturais. Na história, em particular, elas são raras. Este estudo espera preencher lacunas e suscitar novas questões para a área de representação e modelagem do conhecimento na CI.

A pesquisa buscará também contribuir para o entendimento de que as possibilidades abertas nesse cenário transcendem em muito os limites do mundo digital e repercutem diretamente na nossa própria *história em construção,* na maneira como queremos colaborar para a mediação do saber. O ato de conhecer não se reduz a uma apreensão inerte de dados, ou tão somente advém de apreciações meramente lógicas. As pessoas conjugam faculdades cognitivas e perceptivas quando interagem entre si e com os recursos que as cercam, participando da construção do conhecimento sobre a realidade. O ambiente digital e de rede traz infinitas oportunidades de compartilhamento dessa história em construção, com desdobramentos na cultura e no aprendizado. Mas para que esse ambiente seja efetivamente transformador, é fundamental o papel das políticas de difusão e disponibilização dos registros

históricos, bem como a busca pela sua melhor forma de acesso, garantindo a diversidade da qual não podem prescindir. Como no texto de Pierre Lévy:

> "Nem a salvação, nem a perdição residem na técnica. Sempre ambivalentes, as técnicas projetam no mundo material nossas emoções, intenções e projetos. Os instrumentos que construímos nos dão poderes, mas, coletivamente responsáveis, a escolha está em nossas mãos". [LEVY, 1999, p.16]

A dissertação está organizada da seguinte forma:

- Nesta *Introdução* foram apresentados o contexto, pressupostos e questões que orientaram a pesquisa, além dos objetivos que se esperam alcançar.

- No segundo capítulo, *Fundamentos teóricos*, investigam-se os principais conceitos envolvidos que são: i) a Web Semântica**,** seus cenários, padrões e tecnologias; ii) a Representação do Conhecimento sob o viés da modelagem conceitual, construção dos significados, teorias classificatórias e sistemas de conceitos; iii) as Ontologias como instrumento de representação; iv) e finalmente, o Domínio da História, com as especificidades próprias do campo.

- No terceiro capítulo, *Metodologia de pesquisa*, são apresentados os processos pelos quais foram analisados os problemas propostos.

- O quarto capítulo, *Estado da arte e ontologias relacionadas*, reúne ontologias que lidam com a noção de evento sob algum aspecto, discutindo-se os principais requisitos e desafios da modelagem neste quesito.

- No quinto capítulo, *Proposta de modelagem conceitual da ontologia*, é apresentado um estudo para a modelagem conceitual de um esquema inicial de ontologia no domínio da História.

- No sexto e último capítulo, *Considerações finais e possibilidades futuras*, são tecidas considerações sobre os resultados alcançados e contribuições, problemas e limitações encontrados e possibilidades de trabalhos futuros.

2 Fundamentos teóricos

Os principais conceitos envolvidos na pesquisa, como já explicitados na introdução, versam sobre a Web Semântica, a Representação do Conhecimento, Ontologias e o domínio da História, e o objetivo deste capítulo é justamente definir e fixar o entendimento destes conceitos.

Para tanto, seu desenvolvimento pautou-se na seleção de textos cujos autores são considerados seminais em sua área de pesquisa, sendo frequentemente citados pelos demais estudiosos dos temas propostos. Em especial, o movimento de debruçar-se sobre a web semântica e sobre o uso de ontologias em sistemas especialistas, é algo muito atual e pragmático, e por isso a profusão de artigos científicos que vêm sendo publicados nas diversas áreas de aplicação - da Filosofia à Engenharia de Computação, passando pela Lingüística e Inteligência Artificial -, nos últimos anos, cada qual com suas contribuições particulares e olhares específicos. Autores comentadores certamente dão peso às escolhas teóricas e metodológicas, e foram aqui selecionados segundo critérios de produção acadêmica, inserção em instituições e grupos de pesquisa, orientações e capacidade de harmonizar as análises com referências teóricas importantes em seus textos.

Tendo isso em vista, na próxima seção apresentaremos uma visão geral sobre a web semântica e as mudanças previstas na forma como representamos e disponibilizamos os recursos digitais na web; na seção seguinte, um estudo sobre a representação do conhecimento e os pressupostos teóricos subjacentes a esta temática; em seguida, o tema das ontologias e por que considerá-las instrumentos de representação; e finalmente, a História como domínio do conhecimento a ser modelado.

2.1 Web Semântica

Em uma entrevista concedida para o site ConsortiumInfo.org em 2005[5], Tim Berners-Lee aponta que o conceito da web semântica não está na ideia de revolução, mas sim de evolução. Ou seja, estamos a falar da mesma web, porém uma web *estendida*, onde é dada à informação significado capaz de ser lido e interpretado por computadores e por pessoas. Suas sementes foram apresentadas ainda no tempo da proposta original da web, na década de 1990 (BERNERS-LEE, 1998), e desde então muitos estudos nesta seara foram realizados, de maneira interdisciplinar.

Breslin, Davis e Novak (2008) identificam quatro estágios para a web segundo sua capacidade de realizar conexões: o primeiro estágio que trata apenas de conectar informação; o segundo que trata de conectar pessoas – colocando o "eu" na interface de usuário e o "nós" nas redes sociais; o terceiro que diz respeito a representar significados e conectar conhecimento; e o último estágio, futuro, que trata de conectar inteligências colocando pessoas e sistemas raciocinando e se comunicando juntas.

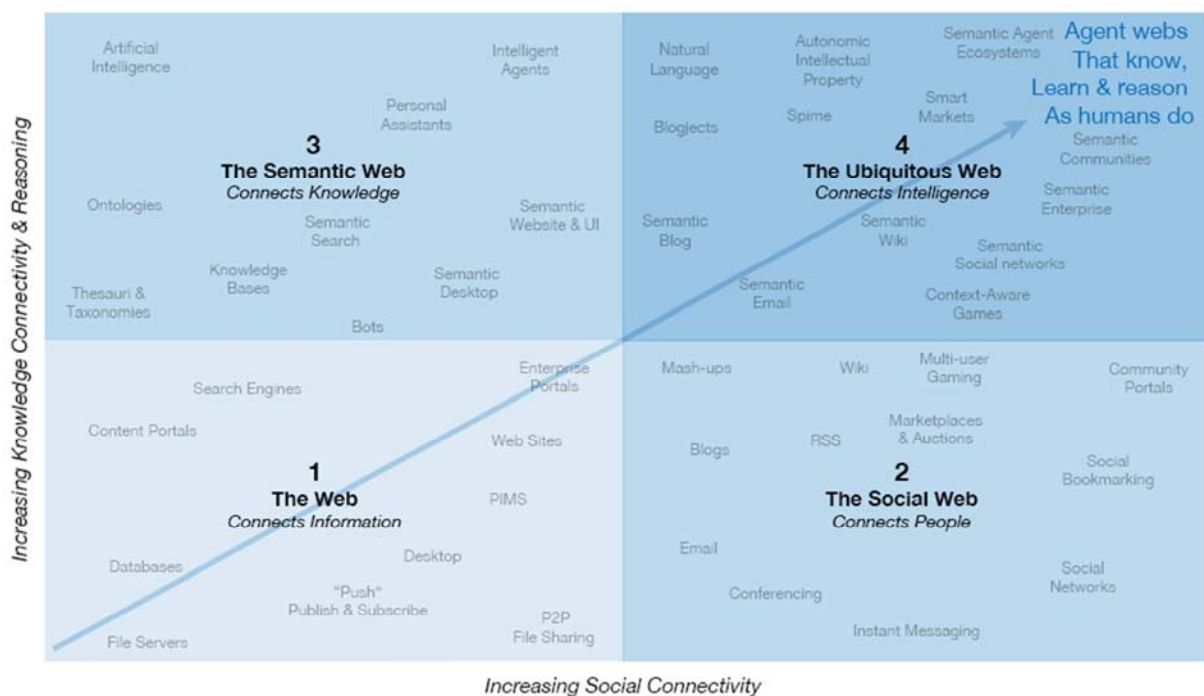



Figura 1 - A evolução da Internet - Fonte: BRESLIN, John; DAVIS, Mills; NOVA, Spivak. Project10X, 2008.

[5] Entrevista concedida para o Consortium Standards Bulletin. Disponível em: http://www.consortiuminfo.org/bulletins/semanticweb.php. Acessado em 25 de janeiro de 2011.

A combinação web + semântica é tratada pelas comunidades científicas a partir de diferentes enfoques. Isso vai muito de acordo com as áreas de pesquisa em que atuam e também a concepção particular que fazem acerca do que seria uma web semanticamente estendida. De um modo geral, as abordagens consideram aspectos que a vêem como uma plataforma de serviços onde interagem agentes inteligentes, como uma tecnologia que adiciona anotações semânticas às informações na web e como um grande repositório de dados conectados. São abordagens que se complementam.

2.1.1 Agentes inteligentes

O artigo "The Semantic Web" publicado na revista *Scientific American* (BERNERS-LEE; HENDLER & OSSILA, 2001) descreve um cenário fictício no qual agentes inteligentes realizam tarefas no lugar de humanos como se fossem assistentes pessoais, interagindo de forma autônoma em busca de soluções adequadas em uma web provida de informações estruturadas. Neste sentido, tais agentes são entidades similares a programas de computador que, acessando bases de conhecimento contendo fatos sobre certo domínio, conseguem realizar análises semânticas sobre conteúdos na web e direcionar tomadas de decisão baseados em elementos de aprendizado sobre este conhecimento (HENDLER, 2001).

Os chamados *Semantic Web Services* (SWS), ou Serviços de Web Semântica, são o ponto de convergência entre a desejável web semântica, os agentes inteligentes e as tecnologias de *web services*. Atualmente os esforços de pesquisa se concentram na identificação mais precisa de requisitos para os sistemas SWS, quais sejam, seu ciclo de vida, arquitetura, tarefas e frameworks desejáveis (BENJAMINS et al, 2011).

O que se espera são serviços capazes de fornecer informações imediatas a partir de requisições específicas, tais como: calcular uma rota para certos destinos ou pontos de interesse; recuperar dados turísticos de uma cidade que se está visitando, obter informações sobre eventos, horários de funcionamento dos locais, previsões do tempo, serviços de emergência próximos, etc. Um bom exemplo retirado de (OBERLE et al, 2006) considera um turista que está visitando Berlim de carro e aciona um *SmartWeb system* a partir de seu smartphone: 'estou com fome. Além disso, quase não há gasolina no carro. Infelizmente não tenho nenhum dinheiro, possuo apenas um cartão de crédito. *SmartWeb*, por favor me ajude.' Neste caso, o sistema interpreta as intenções do usuário e associa as expressões 'fome' com 'restaurante', 'não há gasolina' com 'ir a um posto de gasolina' e 'cartão de crédito' a 'formas

de pagamento'. Para traçar uma rota até o posto adequado mais próximo, a posição do usuário é determinada por meio de recursos integrados de GPS.

Deixando de lado as expectativas de máquinas poderem ser treinadas para se comportarem como humanos, o que se espera da web semântica, apoiada pelos SWS, é o acesso mais inteligente às informações. Neste sentido, seus estudos envolvem linguagens computacionais, lógica e instrumentos de metadados.

2.1.2 Anotações semânticas

Metadados, ou metainformações, podem ser definidos como 'dados que descrevem dados'. Bibliotecas, museus e centros de documentação fazem uso de metadados com propósitos de catalogação e descrição de objetos que integram seus acervos. Também na organização de recursos digitais, providenciam padrões de descrição que facilitam a interoperabilidade entre sistemas de informação. Bibliotecas digitais, por exemplo, utilizam metadados para identificar o tipo de produção, autoria, data de publicação, entre outros, de um determinado recurso. Este conjunto de elementos de descrição, ou *esquema* de metadados, pode ser compartilhado para identificar recursos em outros sistemas de informação, e assim por diante.

Tais modelos de descrição de conteúdo têm sido bastante adotados em ambientes digitais com foco no acesso à informação. Iniciativas como Dublin Core[6], Encoded Archival Description[7] (EAD) e Text Encoding Initiative[8] (TEI) fazem parte do rol de ferramentas desenvolvidas para bibliotecas digitais e representam um enorme avanço na questão de interoperabilidade e troca de informações entre repositórios. No entanto, ainda assim o acesso efetivo aos recursos não se dá plenamente.

Algumas pesquisas mostram que o uso de metadados baseado em ontologias pode ser um caminho promissor. Em um estudo de caso, Weinsten (1998) realizou a conversão de um catálogo bibliográfico codificado no formato MARC para um modelo de ontologia, obtendo uma base de conhecimento com descrições muito mais ricas e expressivas. Isto implica que as consultas passam a ser mapeadas pela estrutura conceitual da ontologia, ou seja, pelos relacionamentos nela embutidos e não somente por atributos de valores.

---

[6] http://dublincore.org
[7] http://www.loc.gov/ead
[8] http://www.tei-c.org

A tecnologia semântica adiciona metadados às informações através do uso de marcações (*tags* em inglês):

```
<pessoa>
   <nome>Getúlio Vargas</nome>
   <cidadeNascimento>São Borja</cidadeNascimento>
   <dataFalecimento>24 de agosto de 1954</dataFalecimento>
</pessoa>
```

As anotações permitem a um computador entender que Getúlio Vargas é o nome de uma pessoa que nasceu na cidade de São Borja, e 24 de agosto de 1954 é sua data de falecimento. De forma bastante rudimentar, o exemplo acima demonstra a essência da tecnologia da web semântica.

Para Greenberg, Sutton & Campbell (2004), a implementação e coleta de metadados é fundamental para o sucesso da web semântica, pois são eles os elementos com os quais os agentes irão interagir.

2.1.3 Linked data - web de relações ricas

Conteúdos semânticos nos permitem mover da web atual de relações simples do tipo ‘linksTo’ para uma web mais rica onde podemos adicionar significado e expressar relações tais como ‘trabalhaPara’, ‘éAutorDe’, ‘localizadoEm’, (MILLER e SWICK, 2003). A figura a seguir ilustra como são referenciados os links na web atual e o que se espera da web semântica:

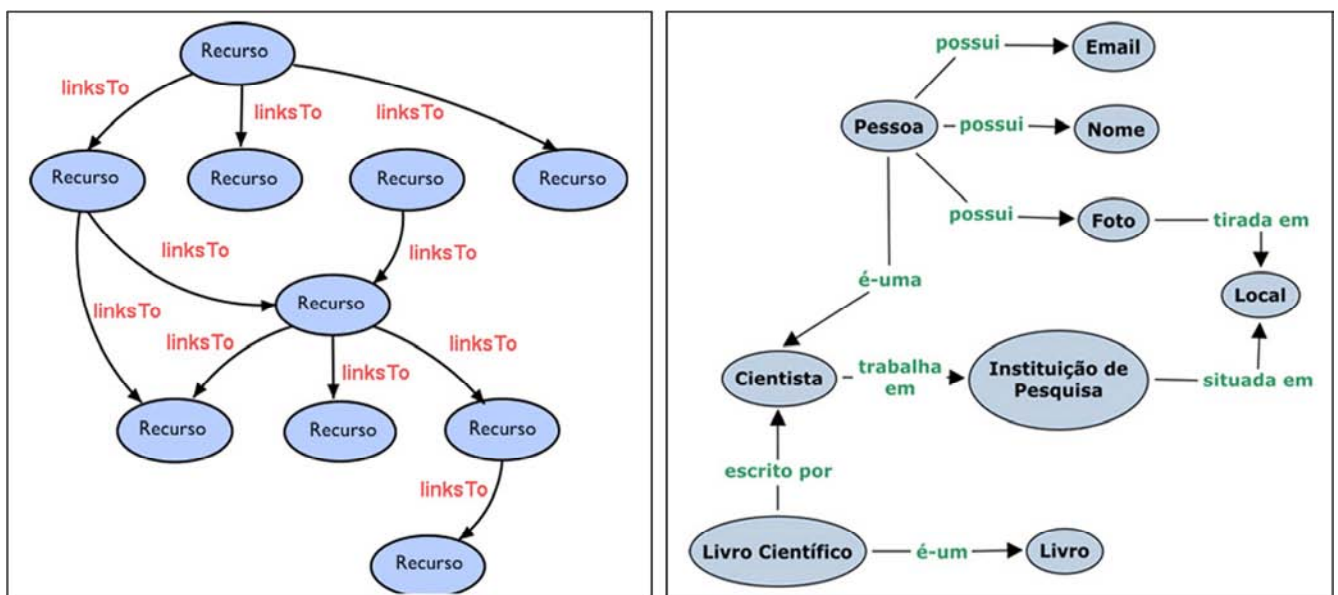



Figura 2 - Estrutura da web atual e o que se espera da web semântica. Adaptado de Miller, 2004 e Guizzardi, 2009.

Esse seria o caminho para a efetiva integração da informação, onde a web se tornaria um verdadeiro banco de dados de "alta resolução" (Web of data). A busca alcançaria não apenas palavras-chave, mas conexões ricas de significado, algo como "aqui estão os pesquisadores que trabalharam na instituição X, que publicaram pelo menos um artigo sobre o assunto Y em periódicos de conceito A, e estabeleceram algum tipo de relação com pesquisadores da instituição Z". Neste mundo ideal das respostas precisas as informações provêem de diversas fontes de dados livres e interoperáveis (*datasets* e ontologias), selecionadas a partir de propriedades e valores que preencham o contexto proposto.

Dados abertos e interligados

O espírito da informação livre comumente associado às licenças de direitos autorais e aos movimentos de software de código-aberto ganha espaço na web semântica a partir do esforço capitaneado principalmente por Tim Berners-Lee em construir uma rede global de dados interligados.

Na apresentação feita para o TED 2009 Conference[9] sobre *Linked Open Data*[10](LOD), ele mostrou como dados abertos e conectados podem constituir uma grande base de conhecimento sobre praticamente tudo. Se dados estão sempre relacionados a outros dados, a possibilidade de se descobrir coisas novas é ilimitada e está diretamente associada ao movimento global de se tornarem disponíveis cada vez mais dados, provenientes de iniciativas particulares ou de governos, empresas e organizações.

No site do Linked Data[11] é possível ter uma ideia da quantidade de *datasets* disponíveis e acessíveis pelos padrões da web semântica:

---

[9] Série de conferências sobre tecnologia, entretenimento e design, promovido pela Sapling Foundation, localizado em http://www.ted.com/.
[10] Disponível em http://blog.ted.com/2009/03/13/tim_berners_lee_web. Acesso em 03/10/2010.
[11] Linking Open Data cloud diagram. http://lod-cloud.net/. Acesso em 03/03/2012.

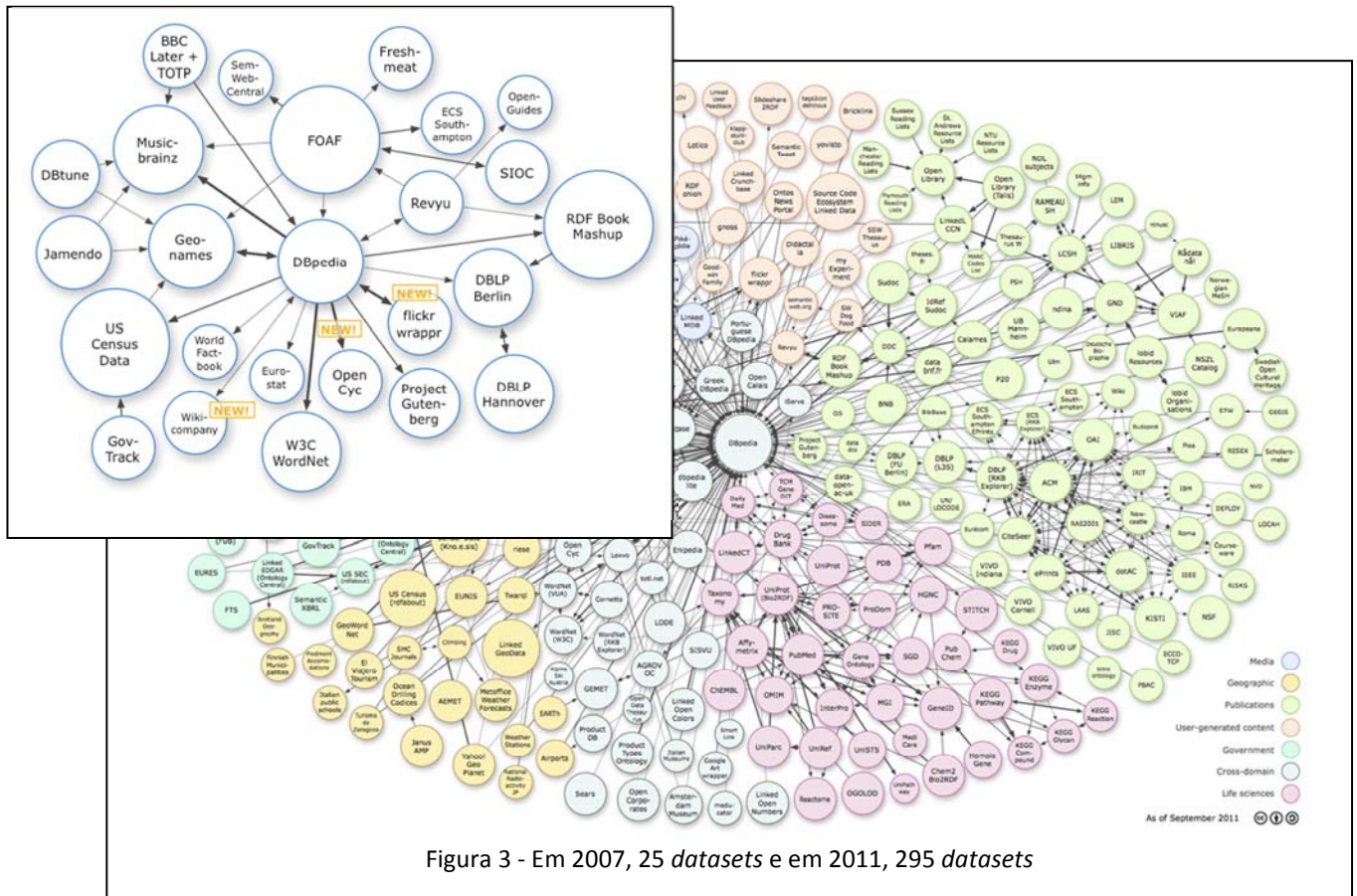



Figura 3 - Em 2007, 25 *datasets* e em 2011, 295 *datasets*

Os *datasets* acima estão agrupados por cores que representam domínios do conhecimento - como ciências naturais, música, geografia, publicações e multi-domínios -, que juntos contabilizam mais de 30 bilhões de assertivas[12] expressas em triplas RDF[13].

Exemplos de *datasets* importantes

O DBpedia[14] detém atualmente cerca de 3,6 milhões de conceitos extraídos de informações estruturadas da Wikipedia, descritos em mais de 1 bilhão de triplas RDF. YAGO[15], que se define como uma grande base de conhecimento semântico sobre pessoas, organizações e lugares, supera o número de dez milhões de entidades e aproximadamente 120 milhões de fatos sobre estas entidades. Sua principal fonte também é a Wikipedia, mas YAGO conta ainda com a WordNet e o GeoNames, outros *datasets*, para estruturar suas

---

[12] Informacao disponível em: http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#Project_Description. 6/6/2012.

[13] Um conjunto de regras e boas práticas para publicar no formato Linked Data foi estabelecido por Tim Berners-Lee e adotado pela W3C. Disponível em: http://linkeddata.org/guides-and-tutorials. Sobre RDF, ver capítulo 5.3.3.

[14] http://dbpedia.org. Acessado em 17/11/2011.

[15] YAGO (*Yet Another Great Ontology*): http://www.mpi-inf.mpg.de/yago-naga/yago. Acessado em 14/04/2012.

informações. A WordNet[16] - assim como suas variantes regionais, incluindo uma iniciativa para o português (RADEMAKER & PAIVA, 2012) - é uma base lexical que se conforma em conjuntos de sinônimos, os *synsets*, que visam a representar conceitos pelas suas relações léxico-semânticas e lógico-conceituais. Muito utilizada em aplicações de processamento de linguagem natural (PLN), é constituída de milhares de termos relativos a substantivos, verbos, adjetivos e advérbios organizados em torno das suas relações conceituais. GeoNames[17] é uma base de conhecimento contendo mais de 10 milhões de nomes geográficos e informações sobre lugares, latitude, longitude, população, divisões administrativas e códigos postais. Freebase[18] é outra base aberta, construída sobre dados coletados externamente ou a partir de contribuições de usuários. Sua estrutura de dados utiliza um modelo de grafo, o que significa que ao invés de usar tabelas e chaves, adota o conceito de nós e links. Dessa forma, consegue-se estabelecer relacionamentos muito mais complexos entre os seus elementos, de uma forma que uma base de dados convencional não permitiria, e está aberta para que usuários entrem com novos objetos e relações nos grafos já criados.

O conceito de *Linked Data* como base para o compartilhamento de dados abertos possibilita a exploração e descoberta de novos conhecimentos. Através da combinação de *datasets* seria possível, por exemplo, avaliar o envolvimento de representantes do povo brasileiro na solução de problemas eminentes em sua região de atuação. No caso da educação, como ilustração, selecionariam-se políticos oriundos de estados com alto índice de analfabetismo, cruzando seus dados com os projetos com os quais estão envolvidos. A taxa de analfabetismo poderia ser extraída de um *dataset* sobre indicadores sociais e as votações de políticos poderiam ser obtidas de outro *dataset* que acompanha as ações políticas destes representantes. E se forem agregadas informações disponíveis em redes sociais sobre opiniões dos usuários, seria possível medir a popularidade deste indivíduo.

---

[16] http://wordnet.princeton.edu. Acessado em 24/10/2010
[17] http://www.geonames.org. Acessado em 14/04/2012.
[18] http://www.freebase.com. Acessado em 24/10/2010.

O Brasil tem uma boa oferta de dados públicos colocado à disposição gratuitamente, porém destes, um número muito ínfimo está em formato aberto seguindo os padrões da W3C[19] (exemplos nesse sentido são os projetos Governo Aberto SP[20] e o LeXML[21]). Para contornar esta dificuldade, iniciativas como as do site Ligado nos Políticos[22] utilizam *web crawlers*[23] para extrair dados de forma metódica e automatizada de bases públicas como

Figura 4 – Ligado nos Políticos: disponibilização de dados abertos

Tribunal Superior Eleitoral (TSE), Senado Federal, Portal da Câmara dos Deputados, Políticos Brasileiros, Ficha Limpa e Transparência Brasil, com o objetivo de construir uma base aberta em formato RDF e conectada a outros *datasets*. São informações sobre nossos políticos, seus dados pessoais, declarações de bens, eleições, mandatos, afastamentos, pronunciamentos, comissões, proposições e ocorrências.

O Grupo de Trabalho sobre Dados Abertos Governamentais da W3C Brasil identificou em seu relatório de 2011[24] as demandas prioritárias de informação a serem supridas pela oferta de dados abertos governamentais, e nele podemos encontrar: dados sobre educação, sobre os eventos da Copa e Olimpíadas no Brasil, de sistema de transporte e trânsito, das

---

[19] A *World Wide Web Consortium* (W3C), fundada por Berners-Lee em 1994, é o principal órgão de padrões da web. O grupo de trabalho *W3C Semantic Web Activity* tem papel de liderança no planejamento de especificações e no desenvolvimento aberto e colaborativo de padrões e tecnologias para a web semântica.
[20] http://www.governoaberto.sp.gov.br
[21] http://www.lexml.gov.br/
[22] http://ligadonospoliticos.com.br
[23] *Web crawler* é um programa de computador que navega pela web e executa tarefas automatizadas, como por exemplo, obtenção de tipos específicos de informação a partir de um conjunto de regras.
[24] http://www.w3c.br/GT/GrupoDadosAbertos. Acesso em 22/11/2011.

pesquisas científicas financiadas com recursos públicos, de desempenho do Legislativo e Judiciário, de geolocalização dos municípios e do sistema jurídico-penal. Devem estar em consonância com as três leis dos dados abertos governamentais divulgado pela W3C[25], quais sejam: 1) se o dado não pode ser encontrado e indexado na web, ele não existe; 2) se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; 3) se algum dispositivo legal não permitir sua reaplicação, ele não é útil.

Estes são os primeiros passos para a web idealizada por Berners-Lee, de dados abertos, acessíveis pelos padrões da computação semântica e conectados por relações que expressam um significado compartilhado. Aos poucos seus modelos e tecnologias parecem emergir de uma experiência de pesquisa para um esforço concreto, produtivo e real, com espaço de aplicação e consumo abrangendo vários setores da sociedade, conforme demonstra o gráfico abaixo, produzido a partir da análise de cerca de 100 empresas americanas e européias provedoras de serviços (BENJAMINS et al, 2011):

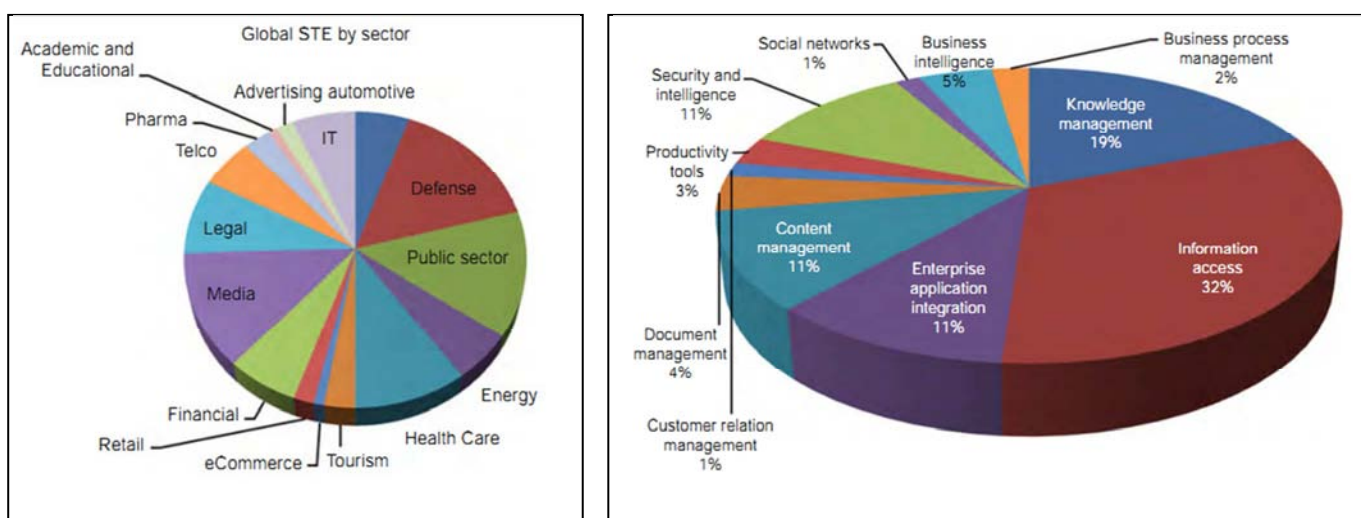



Figura 5 - Setores e serviços que mais consomem aplicações de tecnologia semântica
Fonte: BENJAMINS et al, 2011

Os setores mais representativos no consumo de aplicações com tecnologias semânticas incluem: defesa (15%), governo (14%), mídia (13%) e assistência médica (10%). E de acordo com os autores da pesquisa, essas tecnologias visam aperfeiçoar principalmente o acesso à informação (32%), a gestão do conhecimento (19%), o gerenciamento de conteúdo e a integração de aplicações (11%).

[25] Em http://www.w3c.br/divulgacao/pdf/dados-abertos-governamentais.pdf. Acesso em 22/11/2011.

2.1.4 Arquitetura da web semântica

A arquitetura idealizada pela W3C para a web semântica é ilustrada por um “bolo” de várias camadas, onde cada uma delas suporta ou tem conexão com metadados. Pode ser resumida da seguinte forma (GREENBERG et al, 2003; BERNERS-LEE, 2000; DOMINGUE, FENSEL & HENDLER, 2011):

URI e Unicode: O Identificador Único de Recursos é um metadado que funciona como um ISBN ou documento de identidade no contexto da web, possibilitando a identificação precisa dos recursos. O Unicode é o esquema padronizado de codificação dos caracteres que funciona independentemente de plataforma. Permitirá que o XML contenha a maioria dos caracteres internacionais.

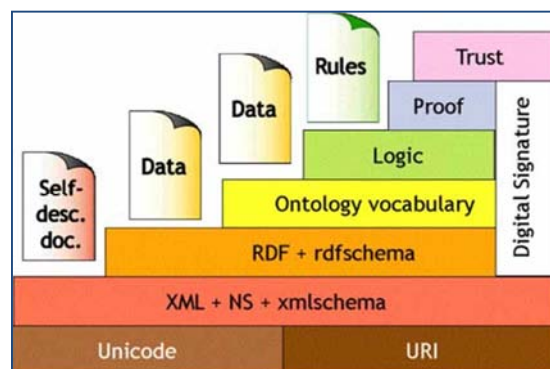



Figura 6 - Arquitetura da Web Semântica
Fonte: GREENBERG et al, 2003

XML, XML Schema e Namespace: O XML (*eXtensible Markup Language*) é uma linguagem de marcação que possui a função específica de estruturar as informações. Ao contrário do HTML (*HiperText Markup Language*) que se preocupa com a exibição dos dados no navegador, o XML define elementos e atributos de metadados. É uma linguagem flexível que permite a criação de marcações personalizadas, conferindo uma estrutura de descrição rica, ainda que arbitrária[26], aos recursos. Scripts e programas fazem uso destas marcações de maneiras sofisticadas, mas é preciso explicitar como cada uma deve ser representada e tratada. É onde entra o *XML Schema* trazendo um conjunto de regras sintáticas que servirão para analisar e validar o conteúdo destes recursos. Já o *namespace*, identificado por um URI, seria a coleção de nomes utilizada para assegurar a interoperabilidade semântica entre vocabulários de metadados.

No exemplo a seguir, o namespace `hist` agrupa alguns metadados que descrevem um personagem histórico (nome, data de nascimento, cargos ocupados), e está associado a um único URI (fictício), `http://pesquisa.dominio.br/historia`:

---

[26] Arbitrária no sentido de que há uma liberdade na criação das tags, porém geralmente isto é feito através de um processo participativo no interior da comunidade de interesse.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<verbete
    xmlns:hist="http://pesquisa.dominio.br/historia">
    <hist:Nome>Getulio Vargas</hist:nome>
    <hist:DataNascimento>19 de abril de 1882</hist:DataNascimento>
    <hist:Cargos>
        <hist:Cargo>
            <hist:Descricao>Ministro da Fazenda</hist:Descricao>
            <hist:InicioMandato>15 de novembro de 1926</hist:InicioMandato>
        </hist:Cargo>
        <hist:Cargo>
            <hist:Descricao>Presidente da República</hist:Descricao>
            <hist:InicioMandato>3 de novembro de 1930</hist:InicioMandato>
        </hist:Cargo>
    </hist:Cargos>
</verbete>
```

Figura 7 - exemplo de código XML. (Fonte: a autora)

RDF e RDF Schema: esta camada possibilita a descrição dos recursos por meio de suas propriedades e valores, provendo um modelo de dados simples, porém poderoso, sobre a qual agentes realizam inferências lógicas baseadas em metadados. RDF é o acrônimo de *Resource Description Framework* e será abordado em detalhes mais a frente.

Ontologia: ontologias representam a artéria central da web semântica e viabilizam a criação de modelos que vão de simples descrições de recursos até esquemas de classificação complexos. Por ser um dos pilares desta pesquisa será discutido com especial atenção em seção própria sobre o tema.

Lógica: podemos considerar a lógica como um modelo matemático do pensamento dedutivo, uma maneira poderosa de derivar novos conhecimentos. A partir das relações ontológicas explicitadas em etapas anteriores, podemos inferir novos fatos a partir de fatos já conhecidos. Esta camada define regras para o tratamento das informações que possibilitará aos agentes realizarem inferências.

Prova e confiança: a prova possibilita a verificação/comprovação da coerência lógica, de modo que os aspectos semânticos das informações estejam descritos de maneira adequada atendendo a todos os requisitos das camadas abaixo dela. A confiança está relacionada à integridade da prova e espera-se garantir a representação correta com certo grau de confiabilidade. A funcionalidade destas duas camadas é altamente dependente da criação de metadados precisos e bem definidos.

Todo esse ferramental é a base para a modelagem de domínios do conhecimento no contexto da web semântica, isto é, para a construção de modelos de representação que irão auxiliar o desenvolvimento de estruturas computáveis e a realização de tarefas sofisticadas de forma automatizada.

2.2 Representação do Conhecimento

Representar conceitos e significados no ambiente digital é a chave para a proposta da web semântica, e por isso sua estreita relação com a questão da representação do conhecimento.

Os termos 'representação do conhecimento' e 'representação da informação' têm sido utilizados em diversos contextos para denominar grupos e linhas de pesquisa, disciplinas e cursos na área da Ciência da Informação. Mas enquanto a representação da informação compreende o conjunto de elementos descritivos de determinado objeto informacional - obtido pelos processos de descrição física e de conteúdo -, a representação do conhecimento se constitui numa estrutura conceitual que representa modelos de mundo, capazes de descrever e fornecer explicações sobre os fenômenos observados; ela é fruto de um processo de análise de domínio e procura refletir uma visão consensual sobre a realidade que se pretende representar. (BRASCHER & CAFÉ, 2008).

Notável por seus trabalhos nos campos da filosofia da natureza e da natureza humana, Immanuel Kant busca em sua *Crítica da razão pura* (1781) respostas às questões sobre como é possível o conhecimento, ou em que bases se fundam a relação com o objeto daquilo que designamos por representação? Ele afirma: "Estamos de posse de determinados conhecimentos *a priori* e mesmo o senso comum nunca deles é destituído" (KANT, 2001, p.63). Para o autor, há o conhecimento anterior de algumas coisas, uma vez que a mente tem que ter estas categorias de forma a poder compreender experiências reais que se apresentam às nossas consciências. Não podemos pensar um objeto sem as categorias, assim como não podemos conhecer um objeto pensado sem as intuições e sensibilidades correspondentes a esses conceitos. Se as categorias são conceitos *a priori*, o entendimento, que é uma faculdade da razão, nos fornece as categorias com as quais podemos operar as sínteses do diverso da experiência (KANT, 2001, p.66).

Seguindo o raciocínio kantiano, um esquema de representação do conhecimento pode então se apoiar numa visão ou pressuposto ontológico que define as categorias do que existe a

priori. No contexto desta pesquisa, interessa aos nossos propósitos identificar algumas destas visões que tiveram influência direta na forma como pensamos e representamos as coisas.

Por cerca de dois mil anos os estudos na área foram influenciados principalmente pelo modelo estabelecido por Aristóteles[27] na sua tentativa de organizar o conhecimento no mundo. Aristóteles apresenta categorias para agrupar seres em gêneros e espécies de acordo com suas substâncias, semelhanças e propriedades e introduz o termo *differentia* para a atribuição de novas propriedades na classificação de espécies de mesmo gênero. Substância identificaria o sujeito, respondendo à pergunta 'o que é?'. Qualidade, quantidade, relação, lugar, tempo, situação, estado e ação seriam propriedades ditas acidentais, porque especificam o sujeito, indicando como ele é, qual seu tamanho, onde ele está, como está, quando está, entre outras coisas (KAULA, 1982; SOWA, 2000; MOREIRA et al, 2004).

Substância	*Sócrates*
Qualidade	*Filósofo*
Quantidade	*Cinco pés de altura*
Relação	*Amigo de Platão*
Lugar	*Na Ágora*
Tempo	*Ao meio-dia*
Posição	*De pé*
Estado	*Mal vestido*
Ação	*Conversando*
Paixão	*Sendo Insultado*

Tabela 1 - Exemplo de instanciação das categorias de Aristóteles

O filósofo vienense Franz Brentano organizou as categorias de Aristóteles na forma de uma árvore, dispondo as propriedades conforme mostrado abaixo (SOWA, 1999):

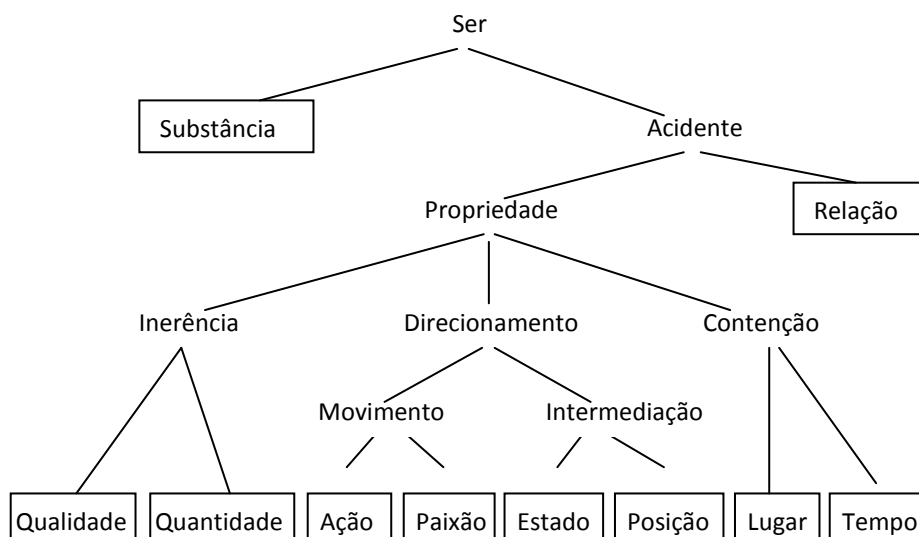



Figura 8 - Árvore de Brentano para as categorias de Aristóteles

[27] Filósofo grego nascido em 384 a.C na antiga Macedônia, um dos maiores pensadores de todos os tempos e considerado o criador do pensamento lógico. Dedicou-se ao estudo dos seres e da natureza e fez da biologia o modelo de suas investigações. Fonte: http://en.wikipedia.org/wiki/Aristotle

Para Fernandez (1976) este modelo aristotélico atendia perfeitamente às necessidades práticas da época e ao universo restrito de então. Mas a partir da invenção da imprensa e das grandes navegações, com as mudanças culturais promovidas pelo homem, além do alargamento das suas relações com o mundo e a descoberta de outros seres e espécies, tornou-se premente a necessidade de critérios mais elaborados para representar e classificar as coisas.

Categorias são, nesse sentido, representações genéricas da experiência – formas de pensar a existência no mundo – e a ordenação da experiência por semelhanças e diferenças lança uma ponte entre a sensibilidade e o entendimento. Inevitavelmente as questões surgem: como se apuram as categorias? Por que se escolhem umas e não outras? Segundo Gil (2000), se não existem critérios, será então de forma empírica, a partir de práticas e de um saber de fato, que se elaboram os quadros categoriais. Aparentemente as categorias não são suscetíveis de uma apreciação para além da utilização específica que delas fazemos em cada caso e segundo a lógica de cada tempo. Na visão de Dahlberg (1978) a formação do conhecimento sobre o mundo acontece exclusivamente através de formas perceptivas dadas a *priori* de espaço e tempo, e através de formas de pensar e raciocinar. Para Alvarenga (2003) o ato de conhecer se utiliza dos sentidos, da emoção, da razão e da linguagem, e a representação do conhecimento compreende um processo cognitivo onde, no percurso, são levantadas as características do ser percebido e a comparação entre estas e as identificadas nos seres já conhecidos. Dos enunciados sobre os seres gera-se o chamado conceito, unidade de conhecimento referente ao ser percebido, componente essencial do conhecimento a ser representado (ALVARENGA, 2003; DAHLBERG, 1978).

A pesquisa em representação do conhecimento tem espaço importante na CI, principalmente através das teorias que fornecem as bases para a construção de modelos conceituais e sistemas de organização do conhecimento (KOS[28]). As categorias aristotélicas nortearam a elaboração da Teoria da Classificação Facetada pelo estudioso indiano Shiyali Ramamrita Ranganathan (1963), e também a Teoria do Conceito por Ingetraut Dahlberg (1978), que veremos com mais detalhes na seção 2.2.3.

Representação do conhecimento também é um ramo da Inteligência Artificial interessada em como o conhecimento pode ser representado simbolicamente e manipulado de forma automatizada por sistemas de dedução e raciocínio (BRACHMAN & LEVESQUE, 2004), os sistemas especialistas. Na Ciência da Computação, além da criação de métodos e

[28] KOS é o acrônimo da expressão em inglês *Knowledge Organization Systems*.

técnicas para captura e explicitação desse conhecimento, a ideia está em que ele possa ser representado em um nível de abstração tal que seja independente e reutilizável ao longo de várias tarefas computacionais (GUARINO, 1997).

Davis, Shrobe e Szolovits (1993) resumem da seguinte forma a utilidade prática atribuída à representação do conhecimento na CI, IA e CC: i) substituir pela abstração coisas que existem na realidade; ii) produzir um conjunto de compromissos ontológicos; iii) servir a uma dada teoria de raciocínio; iv) prover meios para computação eficiente; v) ser um meio de expressão e comunicação para uso humano.

Ao adotar esta abordagem, adentra-se em território anteriormente explorado unicamente por filósofos da ciência e da linguagem, fazendo com que, devido à imposição dessas disciplinas, a área fosse investigada de forma mais pragmática do que quando era um domínio exclusivo da filosofia (GUIZZARDI, 2000).

Almeida e Souza (2011), ao analisarem a noção de ‘semântica’ no âmbito da tecnologia da informação - em particular na visão da web semântica -, propõem um espectro contendo os principais instrumentos usados para representar e organizar o conhecimento. Eles são ordenados de acordo com sua expressividade semântica, separando-se o uso por humanos (linguagem natural) e por máquinas (semântica formal):

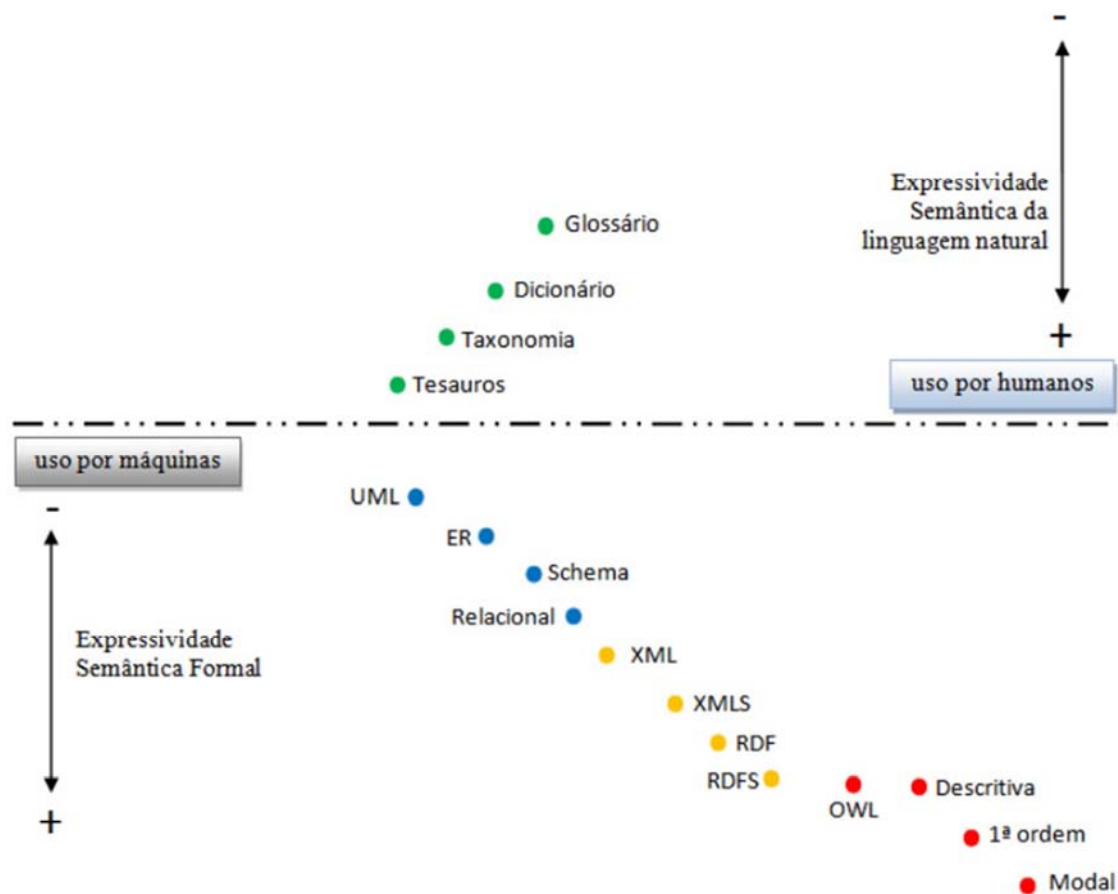



Figura 9 - Expressividade semântica dos instrumentos de representação e organização do conhecimento (Fonte: Almeida e Souza, 2011)

O que é relevante neste espectro, além da expressividade semântica do ponto de vista humano, é aquela que permite a realização de dedução e inferência no âmbito dos sistemas inteligentes. Nesse caso, linguagens formais e lógica são os instrumentos mais expressivos de representação.

A web semântica não diz respeito apenas à disponibilização ou acesso a conteúdos digitais. Exige novas formas de pré-estruturação do conhecimento. Também requer encontrar equivalentes eletrônicos (sistemas de metadados, sistemas de classificação e instrumentos de representação) para todas as suas regras e definições conceituais. O desafio é assegurar que estas transformações reflitam a diversidade do domínio, e não o reduza às limitações de algum algoritmo.

A seguir, apresentaremos alguns estudos que estão intimamente ligados à questão da representação do conhecimento: i) modelos e modelagem; ii) a construção do significado; iii) sistemas de conceitos e teorias classificatórias.

2.2.1 Modelos e modelagem

Buscando aporte na ciência cognitiva, Kaschek (2011) aponta que o conhecimento humano pode ser classificado como sendo de dois tipos: conhecimento declarativo ("saber o que") e conhecimento procedural ("saber como"). O primeiro diz respeito à informação cuja organização tem a forma de séries de fatos conectados e passíveis de descrição; o segundo, em contraste, é o conhecimento implícito, que consiste de capacidades ou habilidades motoras e sensoriais que uma pessoa se engaja para cumprir determinados objetivos, e que por serem tão naturalizadas, são mais difíceis de verbalizar. A modelagem conceitual, para o autor, vem propiciar justamente a conexão entre estes dois conhecimentos, e isso fica evidente na construção dos sistemas especialistas, onde além de ser preciso explicitar o conhecimento declarativo (as assertivas), também são necessárias regras de extração desse conhecimento, ou seja, conhecimento procedural.

Modelagem conceitual como disciplina emergiu nos idos da década de 1970 na ciência da computação, quando se começou a falar de modelos de dados e a abordagem entidade-relacionamento (ER) de Peter Chen se consolidava. Tinha por objetivo descrever o conjunto de elementos usados em um domínio para a concepção de esquemas de conceitos que serviriam à memória que um sistema de software deveria ter desse domínio (COUGO, 1997; WIERINGA, 2011). Nas últimas três décadas, o foco deixou de ser exclusivamente os sistemas de dados, especialmente os bancos de dados, para abranger outras aplicações em metamodelagem, representação do conhecimento, ontologias e metodologias para a área de engenharia do conhecimento. Por se basear em uma notação formal, a modelagem conceitual possui vantagens sobre a linguagem natural porque permite a captura exata das semânticas da aplicação (MYLOPOULOS, 1992).

Modelos

Nos modelos, apenas as propriedades consideradas mais relevantes são representadas. Um mapa histórico, por exemplo, é um modelo de uma região particular da superfície da terra; nele não são mostrados aspectos como a vegetação ou a composição do solo, porque estes não são relevantes para o propósito do mapa, que é o de mostrar eventos, acontecimentos passados ou rotas existentes de um lugar para o outro. Sayão (2001) aponta que a ação de modelar impõe a quem modela uma visão clara e sem ambiguidades do que está sendo modelado, além de exigir uma correta seleção dos elementos do universo do discurso que comporão a visão a ser representada. Assim, modelos permitem que nos concentremos

naquilo que é essencial acerca de uma situação complexa, afastando detalhes não importantes, como ressalta Dodebey (2002):

> “Modelos derivam da necessidade humana de entender a realidade, aparentemente complexa, e são, portanto, representações simplificadas e inteligíveis do mundo, permitindo vislumbrar as características essenciais de um domínio ou campo de estudo”. (DODEBEI, 2002, P.19)

Nos sistemas de informação, modelos comunicam acerca dos processos referentes à aquisição, organização e manipulação da informação e o comportamento do usuário frente a estes processos. De um lado encontra-se o ser humano com sua realidade pessoal, interna e presumidamente idiossincrática, ou seja, com o que chamamos *conhecimento*, e no extremo oposto está o sistema de informação com a sua realidade dependente dos próprios limites da aplicação. Entre esses dois extremos localiza-se o campo de representações, o espaço onde se encontram representações de uma ou de outra realidade. Segundo Burt e Kinnucan (apud SAYÃO, 2002), essas representações tentam criar pontes ou elucidar as estranhezas que se supõe existam entre essas realidades.

Os modelos que estão mais próximos do que seja a representação do usuário humano e do que se passa em sua cabeça em relação ao sistema são denominados modelos cognitivos, ao passo que os modelos que se identificam com o sistema e tentam descrever o que se passa em seu interior são mais conhecidos como modelos de dados. A região intermediária é o lugar dos modelos que interpretam os usuários, o sistema e a interação entre eles, ou seja, dos modelos conceituais. Assim, deve-se ficar clara a distinção que se faz entre a visão individual da realidade – isto é, modelo cognitivo – e a visão percebida por um grupo sobre um dado sistema de informação – isto é, modelo conceitual (BURT e KINNUCAN apud SAYÃO, 2002).

O interesse pelos modelos cognitivos no âmbito da CI tem sido cada vez maior. Segundo Sayão (2002), porque a matéria-prima dos processos cognitivos são “objetos mentais” – conceitos, ideias e conhecimento –, trabalhos nesta área consideram como as pessoas organizam conhecimento, como os conceitos são formados e agrupados na mente humana, e como são categorizados. De acordo com o autor, na área de recuperação da informação se consolidou o consenso de que o sistema homem-computador deve ser visto

como um sistema adaptativo-cognitivo, ou seja, o processamento de informação é sempre intermediado por algum tipo de modelo da realidade.

Enquanto modelos de dados tratam principalmente da descrição de estrutura de dados (relacional, redes, hierárquicos) com vista ao seu acesso e armazenamento, os modelos conceituais têm por objetivo desenvolver uma descrição coerente do significado dos dados, satisfazendo nossas necessidades de conhecimento e conceituação sobre determinado domínio, antecipando ou substituindo a existência de uma realidade qualquer. O que percebemos? Como re-conhecemos a realidade no modelo? O que devemos buscar?

> “... o modelo deve servir como uma fonte unificada de referência, para quando ambiguidades surgirem em discussões sobre este domínio, e como um repositório de conhecimento comum, auxiliando de forma direta a comunicação, o aprendizado e reuso em um nível mais alto de abstração.” (ARANGO, apud GUIZZARDI, 2000, p. 41).

Segundo Campos (2004), existem princípios norteadores que devem auxiliar o modelizador na elaboração de estruturas conceituais em domínios de conhecimento: i) o método de raciocínio utilizado para a organização do conhecimento (indutivo, dedutivo); ii) a unidade que principia a representação do conhecimento (o isolado, o conceito, os objetos que constituem o mundo real); iii) as relações conceituais que dão forma à estrutura do domínio (relações genéricas, partitivas e funcionais). Tais princípios, balizados pelas teorias dos sistemas conceituais, possibilitam ao modelizador construir os pressupostos de onde irá partir para a elaboração de um modelo de representação.

Linguagem de modelagem gráfica

A modelagem é fortemente associada ao uso de uma linguagem gráfica ou de uma metodologia baseada neste tipo de representação: Unified Modeling Language (UML), Agile Modeling, ER Diagrams, ORM Diagrams, Data Flows, Petri Nets, etc, são bons exemplos desta associação. A UML é uma das mais adotadas na área de engenharia de software, fornecendo blocos básicos de construção e regras de combinação destes blocos para modelar qualquer tipo de aplicação (BOOCH et al., 2006). Ela emergiu como uma notação diagramática para a modelagem orientada a objetos, evoluindo para um padrão de linguagem visual que abrange especificação e construção dos objetos e métodos aplicados a todo tipo de

domínio (OMG, 2011). Diferentes diagramas podem ser concebidos para representar as diferentes visões de um sistema:

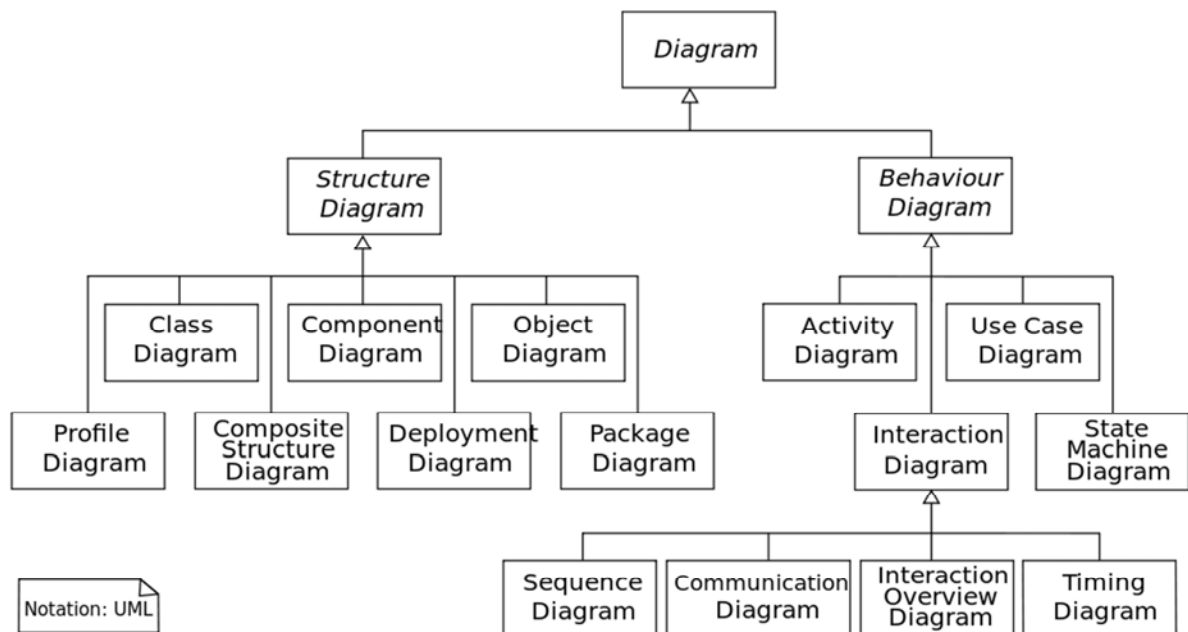



Figura 10 – Visão geral da estrutura de diagramas da UML 2.2
Fonte: Wikipedia (Unified_Modeling_Language) adaptado de (FOWLER, 2003)

A figura acima ilustra a versão 2.2[29] da UML, com seus 14 tipos de diagramas possíveis divididos em duas categorias principais: estrutural (*Structure Diagram*) e de comportamento (*Behaviour Diagram*). Ambas poderiam representar os dois tipos de conhecimento humano que Kaschek (2011) identificou no início desta seção, o declarativo ("saber o que") e o procedural ("saber como") que, combinados, permitem a visualização lógica do desenvolvimento completo de um sistema de informação. Diagramas de classe, em especial, provêem uma visão estática do sistema modelado – onde as operações não estão presentes – e não raro, são usados para a modelagem de vocabulários de sistemas e para a construção de modelos conceituais de domínio, onde todas as entidades são explicitadas (FOWLER, 2003).

[29] Segundo a OMG (*Object Management Group*), um consórcio internacional de empresas que define e ratifica padrões na área de Orientação a Objetos: http://www.uml.org

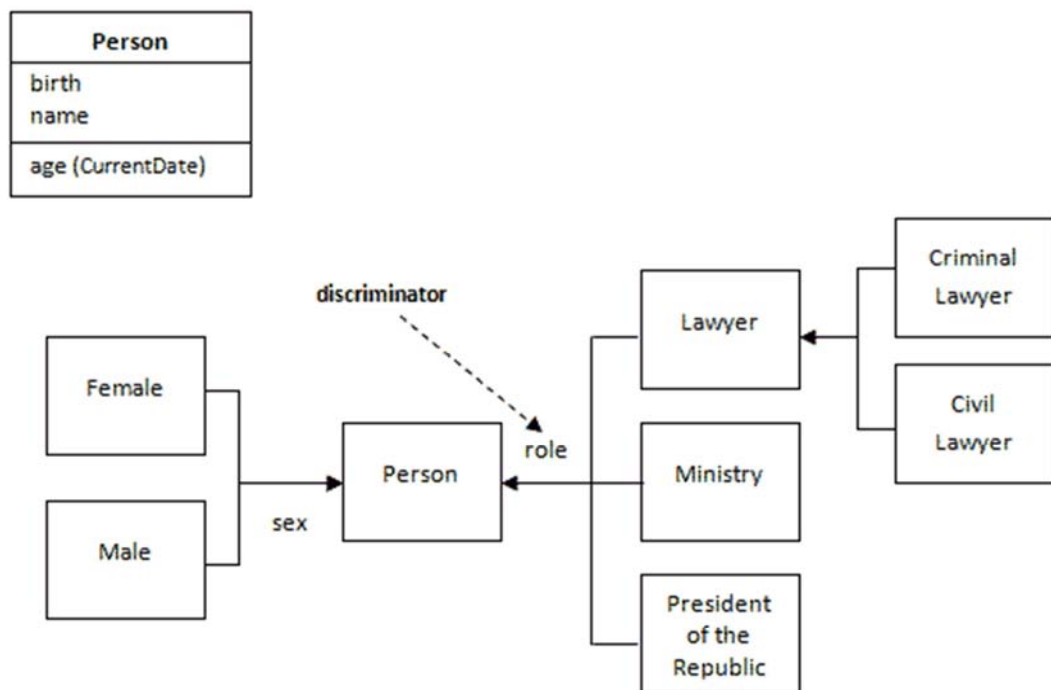



Figura 11 – Acima, à esquerda: exemplo de uma classe em UML. Em seguida, um diagrama de classes. (Fonte: adaptado de FOWLER, 2003)

Na figura acima, as classes, que representam os conceitos identificados no domínio, possuem apenas atributos (nome, data de nascimento e idade) e relacionamentos entre si, tais como os de associação (papéis desempenhados) e de generalização (tipo-de).

A UML, portanto, é uma linguagem de modelagem gráfica que fornece princípios para auxiliar na determinação dos elementos e relações existentes entre os conceitos do domínio que se quer representar, tornando-se expediente de grande importância no contexto da web semântica.

2.2.2 Sobre a construção dos significados

Quando falamos de Semântica, reportamo-nos fundamentalmente ao estudo que tem por objeto a descrição das significações próprias às línguas e sua organização que dá sentido a algo. A significação associa um objeto, um ser, uma coisa, uma noção ou um acontecimento a um signo capaz de evocá-los, e está vinculada, entre outras coisas, às representações mentais que fazemos dos conceitos e das construções nos quais estes participam. (TAMBA-MECZ, 2006; FODOR, 2007).

A questão da representação do conhecimento suscita debates interessantes e multidisciplinares no que diz respeito à construção dos significados. Segundo Morris (1976), Semântica é o ramo da Semiótica que trata das relações entre os signos e os objetos da realidade que eles designam. Para Sowa (2000) o estudo dos significados é indissociável da linguística e principalmente da lógica, que dão base para as semânticas formais, estruturas indispensáveis na modelagem de sistemas computacionais com vistas à representação dos recursos. No contexto da web semântica, Almeida e Souza (2011, p. 36) apontam que "a especificação semântica não corresponde precisamente ao significado dos termos, mas sim ao significado de sentenças de acordo com uma *função interpretação*" sobre expressões sintaticamente bem formadas de um dado modelo de mundo.

Fundamentalmente, para representar mentalmente o mundo que o cerca, o homem precisa articular os seus processos cognitivos utilizando conteúdos generalizados e abstraídos da realidade e guardar tais modelos na memória (MACEDO, 2008). Dessa forma, os conceitos são armazenados em conjuntos ou campos semânticos representados por meio de um modelo mental, a partir do qual ele deriva a experiência e evoca uma realidade ausente.

O processo de significação envolve também a expressão, ou seja, a organização mental dos significados daquilo que se quer dizer para possibilitar a sua exteriorização. Vygotsky (2001, p. 479 apud MACEDO, 2008) afirma que "o significado media o pensamento em sua caminhada rumo à expressão verbal", ou seja, conhecimento é algo que existe na mente das pessoas, mas pode ser compartilhado através de sua representação linguística. A construção do significado é, portanto, não só a compreensão de um conceito, como também a descoberta da sua expressão adequada. Através dos termos como concretização do pensamento é possível representar o mundo e interagir com ele.

A web tradicional, sintática, exibe para o usuário somente as informações na página, ficando ao critério dele estabelecer os relacionamentos de sentido que lhe são válidos. A web semântica, por outro lado, tem como intuito minimizar esse esforço de discriminação, a partir das relações semânticas estabelecidas entre os conceitos que compõem os recursos. Representações de informação têm dois objetivos fundamentais: descrição e discriminação, e devem mapear o "índice intelectual" de um *corpus* particular de informação. Entretanto, segundo Gracioso e Saldanha (2010), as baixas taxas de sucesso percebidas nos SRIs podem ser diretamente atribuídas à indeterminação do significado da linguagem natural nesse

processo de descrição. Para David Blair (2006), a Pragmática[30] ajuda a compreender esse problema, cujo ponto central estaria no estabelecimento de sentido daquilo que é descrito: o que realmente desejamos quando descrevemos o que queremos (sentido)?

Os elementos que interferem nessas ações de descrição e de discriminação de conceitos extrapolam as questões representacionais e atingem configurações comunicacionais. A semântica dos textos é atribuída por pessoas que os indexam manualmente. Isto pode levar a erros infelizes quando, por exemplo, o indexador interpreta equivocadamente o conteúdo de um documento ou não o classifica nas categorias mais relevantes. É comum também encontrar pessoas utilizando os mesmos termos para ideias diferentes ou termos diferentes para as mesmas ideias. Desconsidando os erros crassos de interpretação, percebe-se que a causa de imprecisões e ambiguidades é inerente ao processo de comunicação: em geral, é muito difícil transmitir significados através da linguagem. O significado (semântica) está na mente das pessoas e não na sua representação simbólica.

Frohmann (1990) considera que, para fins de indexação, não basta a compreensão dos processos cognitivos individuais, porque esses não existem isoladamente. Segundo ele, a proposta de Wittgenstein, desenvolvida em seu segundo momento[31], tem grande mérito justamente por demonstrar que a possibilidade de compreensão da forma de pensar do homem está relacionada à sua forma de viver, de se expressar e de se fazer entender pelos membros de sua comunidade. Wittgeinstein rompe com a visão tradicional de que aprender uma língua é dar nomes aos objetos. Para este autor, a linguagem não é uma coisa morta em que cada palavra representa algo de uma vez por todas, ao contrário, é preciso considerar as influências que participam do processo de construção do pensamento socialmente instituído, e, por conseguinte, do significado das coisas. Ou seja, em última análise, se você deseja saber o significado de uma palavra, você deve primeiramente perceber como ela é de fato utilizada e em que contexto. Somente considerando esses princípios será possível representar um recurso de maneira a atribuir-lhe relevância para seus potenciais usuários.

---

[30] Pragmática é o ramo da linguística que estuda a linguagem no contexto de seu uso na comunicação. A pragmática está além da construção da frase, estudado na sintaxe, ou do seu significado, estudado pela semântica: ela estuda essencialmente os objetivos da comunicação, o uso concreto da linguagem. http://pt.wikipedia.org/wiki/Pragmática. Acesso em: 15 de janeiro de 2011.

[31] O pensamento de Wittgenstein, um dos principais filósofos da linguagem, geralmente é dividido em duas fases. A primeira é marcada por estudos que tentam explicar como a linguagem consegue representar o mundo. A segunda avança sobre temas da filosofia da mente ao analisar conceitos como compreensão, intenção e vontade. http://pt.wikipedia.org/wiki/Wittgenstein. Acesso em: 15 de janeiro de 2011.

Nedobity (1989) avalia os diferentes métodos de investigação do significado da linguagem e constata que, enquanto Wüester[32], na Teoria da Terminologia, analisa o significado como sendo construído a partir da relação entre conceitos, Wittgenstein defende a impossibilidade de se prever, estabelecer e reutilizar essas relações conceituais, tendo em vista que elas se constituem exclusivamente em contextos específicos de formas de vida. Ou seja, Wittgenstein busca definir condições gerais de construção da linguagem, enquanto Wüester procura estabelecer leis orientadas para a sistematização de conceitos em contextos específicos. Para Nedobity, a Teoria da Terminologia promoveria condições de sistematização conceitual, mas, dadas as implicações oscilantes de construção de significados, essa precisaria buscar subsídios teóricos mais amplos para expandir suas possibilidades de aplicação, e, mesmo, sua própria concepção, o que poderia ser sustentado por Wittgenstein.

Em *Epistemologia e Ciência da Informação*, Capurro (2003, p.12) afirma que há uma integração entre a perspectiva individualista do paradigma cognitivo e o contexto social nas quais "diferentes comunidades desenvolvem seus critérios de seleção e relevância". Ele argumenta que em um sistema de informação os dados registrados são concebidos por um usuário que desempenha papel ativo em um contexto cultural.

Para Clarinda Lucas (2000) a construção dos índices está indissociavelmente ligada à leitura que os profissionais da informação fazem dos documentos e sua capacidade de sugerir palavras, as instituições em que atuam e o uso que essas instituições fazem da memória. Diz a autora:

> "O método de leitura do bibliotecário não garante a equivalência de sentido entre o texto-fonte e a sua representação. As operações de análise e síntese para fins de análise documentária estão sujeitas a muitos outros fatores para além daqueles que uma metodologia possa disciplinar, administrar, conter." (LUCAS, 2000, p. 63)

Quais seriam as implicações destas questões no domínio da História? De acordo com Hjørland (1999) toda informação deve ser analisada a partir de princípios situacional, pragmático e do domínio específico, a fim de que possa ser adequadamente representada nos sistemas de informação. Como em todas as áreas de conhecimento, os historiadores também

---

[32] Eugen Wüester, engenheiro austríaco que na década de 1930 fundou os princípios da Teoria Geral da Terminologia.

se agrupam em torno de correntes de pensamento, objetos de estudo e métodos de trabalho, formando grupos muitas vezes divergentes que debatem e defendem seus pontos de vista; isso faz parte do processo de validação do conhecimento histórico e impulsiona seu avanço. Ao longo do último século, algumas abordagens influenciaram a forma de "fazer história" no Brasil[33], e este pode ser o ponto de partida para reflexões metodológicas sobre este domínio e determinante no exercício de criação de um modelo conceitual de representação.

A web semântica não visa resolver os problemas de indeterminação do significado no processo de descrição da informação. Ou seja, ao se propor a construção de modelos conceituais não se estabelece o fim das diferenças conceituais ou de significação, até porque, como já mencionado, eles são fruto da cognição particular de cada um e envolvem motivações, crenças, modos de raciocínio e tantos outros estados mentais necessários ao processo de comunicação. Mas um dos pressupostos é que se trata de modelar conhecimento cuja conceitualização será compartilhada e *consensual* para um determinado universo do discurso, e por isso deve refletir um aspecto da realidade observada. Como aponta Almeida e Souza (2011): na semântica formal, "saber o significado de uma sentença equivale a conhecer suas condições-verdade, o que não é o mesmo que saber o seu valor-verdade, ou seja, se o fato é verdadeiro ou não". Dessa forma, através de relações conceituais bem formadas e instrumentos adequados de representação, é possível estabelecer pontes que atravessam as múltiplas e diversas interpretações atribuídas para o dado domínio.

2.2.3 Sistemas de conceitos e teorias classificatórias

Para representar certo fenômeno ou uma dada realidade, é necessário concentrar a atenção em um número limitado de conceitos, que sejam suficientes e relevantes.

Um conceito é uma ideia ou noção que aplicamos a elementos do domínio, e que possui dois aspectos importantes - intensão e extensão. A intensão do conceito é seu significado ou sua completa definição, enquanto que a extensão é o conjunto de elementos de um determinado universo para os quais o conceito se aplica. Por exemplo, o conceito Mortal pode ser definido como 'tudo que morre'. A extensão desse conceito seria então o conjunto de coisas de um determinado universo que possuem a característica de morrer (GUIZZARDI, 2000). Como observado por Campos (2001, p. 72), "extensão e intensão são formas de apreensão e identificação do conceito, e influenciam a elaboração de sua definição".

[33] Por exemplo, abordagens mais positivistas com olhar descritivo e factual; ou abordagens que problematizam a história sob diversos vieses.

Dahlberg (1978) define a formação do conceito como a reunião e compilação de enunciados verdadeiros sobre determinado objeto, fixadas por um símbolo linguístico. Cada enunciado apresenta um atributo predicável do objeto que, no nível do conceito, se chama *característica*. Por exemplo, se tomarmos o conceito "Instituição", sobre ele é possível formular os seguintes enunciados:

- é constituída de um grupo de pessoas;
- que trabalham com determinada finalidade;
- possuindo administração comum;
- localizada em determinado lugar;
- durante determinado tempo, etc.

O conjunto de tais enunciados constitui o conceito "Instituição". Assim como existem diferentes espécies de objetos e de conceitos, existem também diferentes espécies de características dos conceitos: características essenciais (necessárias) e características acidentais (adicionais ou possíveis). A determinação das características e as associações com outros conceitos é o que possibilitará o correto posicionamento do conceito, ou seja, sua ordenação classificatória, em um sistema de conceitos (DAHLBERG, 1978).

Quando falamos em classificação, a ideia de ‘arrumação’ nos vem à mente conjugada com a percepção de que se trata de uma operação naturalizada em nosso cotidiano e essencial para que possamos viver com um mínimo de regras e ordem. Classificar significa agrupar coisas que têm uma característica ou uma qualidade em comum e, por conseguinte, distinguir e separar aquelas que não possuem essa mesma característica ou qualidade. (FERNANDES, 1976; RIBEIRO, 1997; KAULA, 1982). Cada grupo forma, assim, uma classe.

No âmbito da CI, falar sobre classificação é reconhecer nessa prática o processo de organização das fontes de informação. Segundo Almeida e Bax (2003) isso se dá de diversas maneiras: a partir de listas de termos, em glossários e dicionários; por ordenação, através de cabeçalhos de assunto e taxonomias; ou a partir de conceitos e relacionamentos, utilizando tesauros, redes semânticas e ontologias.

Bibliotecas tradicionais utilizam cabeçalhos de assunto para a catalogação de seus documentos, e como o nome indica, o assunto é a unidade primária de informação e seu único ponto de acesso. Com a entrada de sistemas de informação mais evoluídos, a unidade de representação temática deixou de ser apenas o assunto, podendo ser a reunião de conceitos.

Segundo Campos et al (2002) o emprego de palavras-chave foi a primeira iniciativa quando se pensou em romper com o tradicional cabeçalho de assunto e desenvolver tecnologias de múltiplo acesso. No entanto, logo se percebeu que palavras isoladas (descritores) fora de contexto são inadequadas em alguns casos. Passou-se então das palavras-chave para um instrumento de vocabulário controlado de termos relacionados semântica e logicamente, o tesauro. Nele, as relações entre os termos são explicitadas através de indicadores padronizados.

Os tesauros apresentam-se como estruturas consistentes de organização de domínios do conhecimento podendo servir para representar os assuntos nos documentos e nas solicitações de buscas em sistemas. Enquanto as taxonomias abrigam somente relacionamentos de gênero/espécie e todo/parte na classificação dos itens, os tesauros admitem um número maior de relações: associativas, de equivalência, de oposição, de efeito (GOMES, 1990; MOREIRA et al, 2004; DODEBEI, 2002).

Na década de 1960, Ranganathan apresenta uma série de princípios para a organização dos conceitos de um domínio dentro de classes, as quais são estruturadas de acordo com um recorte de categorias fundamentais (CAMPOS, 2001). A metodologia desenvolvida por ele leva em consideração as características dos conceitos como base de divisão para a formação das subclasses, cadeias e renques.

Mais tarde, o método de raciocínio analítico-sintético de Dalhberg (1978) é apresentado para organizar os conceitos em grandes categorias a partir da análise da sua definição - pensamento analítico - e da descoberta de características comuns que revelam categorias para agrupar esses conceitos - pensamento sintético. (CAMPOS, 2011, p.120).

Ambas as teorias influenciaram enormemente a forma de construção de tesauros nos anos seguintes, dando origem a esquemas como o *Thesaurofacet,* baseado no uso de categorias para a organização dos conceitos e posicionamento mais adequado dos termos associados aos conceitos, e os tesauros terminológicos (ou tesauros-com-base-em-conceitos), que conjuga princípios defendidos por Dahlberg e a teoria da classificação de Ranganathan, abordada adiante.

Com relação à categorização, Campos e Gomes (2008) reforçam que este

> "...é um processo que requer pensar o domínio de forma dedutiva, ou seja, determinar as classes de maior abrangência dentro da temática escolhida. Na verdade, aplicar a categorização é analisar o domínio a partir de recortes conceituais que permitem determinar a identidade dos conceitos (categorias) que fazem parte deste domínio" (CAMPOS e GOMES, 2008, p. 5).

A arte de elaborar sistemas, com alguma ideia intuitiva sobre divisões e prioridades no arranjo através de hierarquias e subordinações está no bojo do que se constituiu a Teoria da Classificação, calcada nos estudos de Ranganathan. Dahlberg o considera "pai" da moderna teoria da classificação. Antes dele, sistemas de classificação como a Classificação Decimal de Dewey (CDD) e Classificação Decimal Universal (CDU) já eram usados para a organização do conhecimento bibliográfico. A escolha de números decimais para suas categorias permitia ao sistema ser ao mesmo tempo puramente numérico e infinitamente hierárquico, contendo alguns mecanismos de uma classificação facetada, ao combinar elementos de diferentes partes da estrutura para construir um número representando o assunto.

Mas segundo Dahlberg, o sistema de Ranganathan diferia dos outros, sobretudo pelo fato de que ele não usava classes pré-estabelecidas, mas as criava somente no momento em que um documento era analisado, sintetizando-as a partir das fórmulas de facetas ligadas às disciplinas (DAHLBERG, 1976).

Ranganathan desenvolve sua teoria de classificação baseada na fragmentação de um assunto completo em facetas, ou seja, em grupos de classes reunidas por um mesmo princípio de divisão. Algumas categorias são eleitas como sendo fundamentais: Personalidade, Matéria, Energia, Espaço, Tempo, ou simplesmente PMEST. *Matéria* seria o material físico (ou tudo que se relaciona às características das "personalidades") do qual um assunto pode ser composto. *Energia* seria a ação, ou processo que ocorre em torno de um assunto. *Espaço,* o componente geográfico da localização de um assunto. *Tempo,* o período associado a um assunto. E *Personalidade,* a categoria considerada por Ranganathan como indefinível, ou seja, se certa manifestação for facilmente determinada como *não sendo* espaço, energia ou matéria, ela é considerada uma manifestação de Personalidade.

Mas para Vickery (1960) o campo da ciência e tecnologia já provou que é necessário ter uma lista maior de categorias básicas, e propõe outras além: substância (ou produto), organismo, constituinte, estrutura, forma, propriedade, objeto da ação, ação, operação, processo, agente, espaço e tempo. Sobretudo, diferentes campos de especialização demandarão categorias distintas e não necessariamente exaustivas (KUMAR, 1981).

Segundo Campos (2001), a grande problemática dos esquemas de classificação sempre foi a adequação dos assuntos tratados à estrutura classificatória existente nos esquemas. Até a década de 1930, os esquemas eram incapazes de acompanhar o desenvolvimento do que Ranganathan chamou de “Universo de assuntos dinâmico, infinito, multidimensional, multidirecional e sempre turbulento” (RANGANATHAN, 1967, apud CAMPOS, 2001). Ranganathan evidencia que, com o progresso da Ciência, os livros não mais abordam apenas um único aspecto de um assunto, mas têm um grau de complexidade que, na maioria das vezes, reúnem conhecimento de áreas diversas. Seria necessária uma estrutura classificatória que previsse tal complexidade. Ele apresenta então a Teoria Dinâmica, onde a base classificatória é regida pela hospitalidade, possibilitando a inclusão de novos assuntos sempre que demandado. Além disso, o assunto não está pronto no esquema, mas é construído no momento da análise do documento (CAMPOS, 2001).

Para os sistemas de classificação, Ranganathan distingue três níveis: o nível das ideias e conceitos (plano da ideia); o nível da expressão verbal dos conceitos (plano verbal); e o nível da fixação dos conceitos em formas abstratas, tais como sinais, letras, números (plano notacional). A percepção dos três planos de trabalho levou ao estabelecimento de uma série de cânones e conceitos gerais.

O plano da análise dos conceitos, para Ranganathan, é o plano superior a partir do qual é possível introduzir um mundo inteligível no trabalho da classificação, manifestado através do plano verbal e do notacional. No plano verbal temos a comunicação da ideia ou conceito, mediada pela linguagem, atentando-se para as ocorrências de homonímia, sinonímia e outros. E no plano notacional é onde se darão a representação dos assuntos e a manipulação do arranjo dos documentos.

Sobre os processos de análise facetada, além de Ranganathan, o Classification Research Group[34] (CRG) sediado em Londres, também contribuiu ricamente com novos princípios, tornando-se importante referencial teórico na área. Seus estudos são conduzidos por um núcleo de pesquisadores dispersos pelo mundo, envolvidos na criação e fundamentação de vários esquemas de classificação para a organização do conhecimento.

Classificação é então um mecanismo de abstração através do qual percebemos a existência de elementos na realidade com algumas características em comum e atribuímos um nome (conceito) a esse conjunto (classe) de elementos. Ela resulta do pensamento e da ação humana e social, contendo, por isso, uma margem variável de imprecisão e de representação subjetiva.

A Terminologia possui estreita ligação com a classificação. Em seus estudos, Wüester[35] aborda as diferenças entre sistemas de conceitos e tabelas de classificação, enfatizando o que possuem de semelhante à luz da Teoria Geral da Terminologia (TGT). Ele estabelece a prioridade do conceito em detrimento do termo, considerando o domínio dos conceitos e o dos termos como dois domínios independentes: o termo seria a unidade de comunicação que representa/designa o conceito e o conceito seria o conteúdo do termo, isto é, a unidade de significação que deve ser observada em seu ambiente natural de ocorrência, ou seja, em um dado universo de discurso (WÜESTER, 1971 apud CAMPOS, 2001). Na terminologia, o termo como representante do conceito é normalmente designado por um especialista, ou grupo de especialistas, e determinado de forma prescritiva, segundo a utilidade, sempre dependente do sistema de conceitos. Um dos princípios estabelecidos pela TGT é a precisão do conceito, o que leva, via normalização, à eliminação da ambiguidade e a busca da univocidade, ou seja, a correspondência única entre significante e significado (ou entre denominação e conceito) em um dado contexto (WÜESTER, 1981).

Relações conceituais

Autores como Kietz, Maedache e Volz (2000, apud SALES, CAMPOS & GOMES, 2008) afirmam que a determinação de relações conceituais consistentes é útil à elaboração de

[34] Alguns de seus pesquisadores mais conhecidos incluem J. Farradane, D. Foskett, J. Mills, A.Mills, e o próprio B. Vickery.

[35] O engenheiro austríaco Eugen Wuester, após a experiência de organizar uma terminologia de eletrotécnica com o objetivo de garantir comunicação precisa nesse campo da ciência, desenvolve uma série de princípios cunhados sob o que ele chamou de “estudo científico geral da terminologia”, e que passou a ser conhecida como Teoria Geral da Terminologia (TGT).

ontologias, na medida em que estas relações garantem a consistência na adoção de algoritmos baseados em regras de associação.

Em Gomes (1990) encontramos descritos os seguintes tipos de relacionamentos passíveis de existirem em um sistema de conceitos:

- Relacionamentos lógicos - obtidos da comparação de dois conceitos. Podem se dividir em: i) genérico-específico, que permite identificar as características de pertencimento a uma classe e assim, estruturar hierarquicamente o sistema; ii) relacionamento analítico, que estabelece as relações associativas entre os termos; iii) relacionamento de oposição, que indica os termos opostos;

- Relacionamentos ontológicos – obtidos a partir das relações indiretas entre conceitos. Reune as seguintes relações: i) relações partitivas, que é a relação entre o todo e suas partes; ii) relacionamentos de sucessão, que traz noções de ordenamento no tempo; iii) relacionamentos material-produto, que implicam em mostrar os diferentes estágios na produção de bens.

- Relacionamentos de efeito – constituídos pelos relacionamentos: i) de causalidade ou causa-efeito; ii) de instrumentalidade, relativo ao instrumento e sua ação; iii) de descendência, que guardam entre si uma relação genealógica ou ontogenética.

Como bem apontado em Sales, Campos e Gomes (2008), nas ontologias o conjunto de relações é mais rico do que nas tabelas de classificação bibliográfica ou nos tesauros, o que permite maior estrutura representativa do conhecimento registrado em um discurso, passível de ser manipulada pelo computador. As autoras sistematizam as relações para ontologias em duas classes: Relações Categoriais e Relações Formais, identificando em cada uma delas as seguintes subdivisões: relações genéricas, partitivas e funcionais. A seguir, a sistematização destas relações:

- Relações genéricas: segundo as autoras, as relações genéricas se dão a partir do princípio lógico de abstração, respondendo à pergunta “é um?” Para cada par de relação categorial, foi encontrada uma relação formal:

QUADRO 1 – Sistematização das relações genéricas

Relações Categoriais Genéricas	Relações Formais Genéricas
Gênero-espécie	Tem-tipo
Espécie-gênero	É_um
Espécie-instância	É_instanciado_por
Instância-espécie	É_instância_de

Figura 12 – Relações conceituais genéricas para ontologias
Fonte: Sales, Campos & Gomes (2008)

- Relações partitivas: são aquelas que revelam a relação entre o todo e suas partes, e se dão sempre no interior de uma mesma categoria:

QUADRO 2 – Sistematização das relações partitivas

Relações Categoriais Partitivas	Relações Formais Partitivas
• Objeto-constituinte • Constituinte-objeto	tem_constituinte constituinte_tem_todo
• Objeto-unidade • Unidade-objeto	tem_unidade unidade_tem_todo
• Coleção-elemento • Elemento-coleção	tem_elemento elemento_tem_coleção
• Massa-porção • Porção-massa	tem_porção porção_tem_massa
• Objeto-material • Material-objeto	tem-material material_tem_objeto
• Área-lugar • Lugar-área	tem_lugar lugar_está_em
• Zona-região • Região-zona	tem_lugar lugar_está_em
• Caráter-atividade • Atividade-caráter	tem_etapa tem_processo
• Processo continuo - ato • Ato - processo contínuo	tem_etapa tem_processo
• Processo descontínuo-fase • Fase-processo descontínuo	tem_etapa tem_processo
• Grupo-membro • Membro-grupo	tem_membro membro_tem
• Classe-membro • Membro-classe	tem_membro membro_tem
• Substância-particular • Particular-substância	tem_substância substância_tem
• Conjunto-subconjunto • Subconjunto-conjunto	tem_subconjunto subconjuunto_tem_
• Conjunto-elemento • Elemento-conjunto	tem_elemento tem_conjunto
• Todo/sub-parte temporal • Sub_parte temporal - todo	tem_subparte tem_todo
• Ciência - objeto de estudo • Objeto de estudo - ciência	tem_ramo ramo_tem

Figura 13 – Relações conceituais partitivas para ontologias
Fonte: Sales, Campos & Gomes (2008)

- Relações funcionais: são as que revelam a relação de objeto com o mundo ou a função de um objeto em um determinado contexto. Podem se manifestar entre categorias e intra-categorias. No caso deste estudo pelas autoras (SALES, CAMPOS & GOMES, 2008), a base para a sistematização foram as categorias ranganathianas PMEST (personalidade, matéria, energia/processo, espaço e tempo):

QUADRO 3 – Sistematização das Relações Funcionais

Relações Categorias Funcionais	Relações Formais Funcionais
Personalidade – Personalidade	• é_produto_de • é_instrumento_de • transforma-se_em • deriva_de • derivado_de • interage_com • manifestação_de • analisa • mede • cria • localizado_em • adjacente_a • circunjacente_a • transversal_a
Processo – processo	• método_de • interage_com • é_sub_evento_de • é_fase_de • é_etapa_de • causa • impede
Processo – Personalidade	• causa • afeta • rompe • destrói • resulta_em • localizado_em
Espaço-Espaço	• adjacente_a
Processo – Espaço	• ocorre_em
Medida – Personalidade	• é_medida_de
Medida – Processo	• é_medida_de
Personalidade – Processo	• usado_em • afeta • realiza/produz • é_usado_para • analisa • mede • diagnostica
Personalidade – Propriedade	• tem_propriedade
Propriedade – Personalidade	• é_propriedade_de • é_contraagente_de
Propriedade-processo	• é_propriedade_de

Figura 14 – Relações conceituais funcionais para ontologias
Fonte: Sales, Campos e Gomes (2008)

Assim, esta seção procurou mostrar que a normalização terminológica, a definição precisa dos conceitos, seus recortes classificatórios e o estabelecimento das relações conceituais, constituem um primeiro passo para a construção consistente de vocabulários e ontologias de domínio.

2.3 Ontologias como instrumento de representação

Ontologia como um ramo da filosofia é a ciência que trata do ser enquanto ser, isto é, sua natureza, existência e questões metafísicas em geral. Smith (2002) aponta que a disciplina não se propõe a buscar explicações, mas sim a descrever em termos de classificação as entidades existentes no universo, incluindo os fenômenos e relações ligados a estas entidades.

Desde o início dos anos 1990, ontologias tem sido objeto de estudos no interior de alguns campos da Inteligência Artificial, como engenharia do conhecimento, processamento de linguagem natural e representação do conhecimento (STUDER, BENJAMINS, FENSEL, 1998). A partir da década seguinte, outras disciplinas também tomaram para si este estudo visando aplicações nas áreas de integração de sistemas, gestão do conhecimento, recuperação da informação e descoberta do conhecimento.

Das inúmeras definições concebidas desde então, uma das mais utilizadas é a que Gruber (1993) cunhou como sendo 'uma especificação explícita de uma conceitualização'[36]. Borst (1997) a adaptou levemente acrescentando: 'uma especificação *formal* de uma conceitualização *compartilhada'*. Studer et al (1998) reuniu ambas as definições, explicando: trata-se de uma especificação *formal* porque é legível para computadores; *explícita*, pois os conceitos, propriedades, relações, restrições e vocabulário estão explicitamente definidos em linguagem declarativa; *compartilhada* porque o conhecimento é consensual; e *conceitualização*, porque se trata de um modelo abstrato de algo que se deseja representar.

Dentre outras interpretações, há a que apresenta ontologia como um artefato de software contendo conhecimento codificado em formato processável por máquina (HAFNER apud GUIZZARDI, 2000; NOY & MCGUINESS, 2001), como um tipo especial de base de conhecimento (GUARINO & GIARETTA, 1995; GRIMM et al, 2011) e também como uma teoria lógica, isto é, um conjunto de axiomas expressos em linguagem lógica (GUARINO, 1998; RADEMAKER, 2012). Uma ontologia não descreve apenas conhecimento imediato, isto é, conhecimento factual obtido diretamente da observação do domínio, mas também conhecimento derivado de inferências sobre o conhecimento explícito.

[36] "Conceitualização" não existe no português, porém o termo mais próximo que seria "conceituação", é definido no Dicionário Houaiss como "ato ou efeito de conceituar", o que não corresponde ao seu uso aplicado na área da representação do conhecimento. "Conceitualização" é amplamente adotado para se referir a uma visão abstrata e simplificada do mundo que se deseja representar, considerando as entidades e relacionamentos nele presentes.

Um modelo que utiliza ontologias, portanto, não é somente uma hierarquia de termos, mas uma infra-estrutura teórica que versa sobre o domínio em questão (GUIZZARDI, 2000) e cujo compromisso limita-se ao seu universo do discurso, aos objetos que nele podem ser representados através de linguagem lógica declarativa (GRUBER, 1993).

Ontologias podem ser classificadas sob diversos enfoques. Quanto ao grau de formalidade, Uschold (1996) as define como: i) altamente informais, quando são expressas livremente em linguagem natural; ii) estruturadas informais, quando em linguagem natural, de forma restrita e estruturada; iii) semiformais, quando em uma linguagem artificial, definida formalmente; iv) rigorosamente formais, quando expressas através de semântica formal, teoremas e provas. Quanto à generalidade daquilo que se pretende descrever e representar, Guarino (1998) as classifica como sendo de topo, de tarefas, de domínio e de aplicação. A primeira é independente de domínio e as outras concebidas em função de contextos e propósitos específicos. Além disso, quanto à expressividade, é comum os autores distinguirem ontologias construídas sob um alto grau de formalismo e rigor semântico, com restrições e axiomas bem definidos (*heavyweight/full-fledged ontologies*) daquelas que apresentam apenas um pequeno conjunto de primitivas básicas: conceitos, propriedades e relacionamentos hierárquicos do tipo é-um (*lightweight ontologies)* (MASOLO et al, 2003; CORCHO et al, 2003, GIUNCHIGLIA, 2007; GRIMM et al, 2011).

Os primeiros trabalhos sobre ontologias em CI – em especial na apropriação de ontologias *lightweight* - possuíam forte vínculo com as linguagens documentárias, chegando mesmo a haver pouca discriminação na terminologia do uso, particularmente junto aos tesauros, pelas suas semelhanças de estrutura – ambas organizam conceitos e relações de um domínio – e finalidade – para fins de representação e recuperação da informação (MOREIRA, 2003). No entanto, em termos de linguagem e nível de formalização existe uma diferença entre os dois instrumentos, já que tesauros cumprem o propósito de comunicação entre o usuário e as linguagens documentárias, e as ontologias o de descrição de objetos digitais para inferências computacionais.

Apresentação gráfica

Principalmente pelo ramo da engenharia do conhecimento, ontologias são frequentemente apresentadas sob a forma gráfica, como na figura 15, onde nós (*nodes*) e arcos (*arcs*) representam, respectivamente, conceitos e relacionamentos. Apesar de não servirem à

modelagem mais complexa (com emprego de axiomas, por exemplo) nem serem passíveis de processamento por máquina, este tipo de visualização é bastante adequado para comunicar hierarquias taxonômicas simples de um dado domínio e as relações atribuídas aos conceitos.

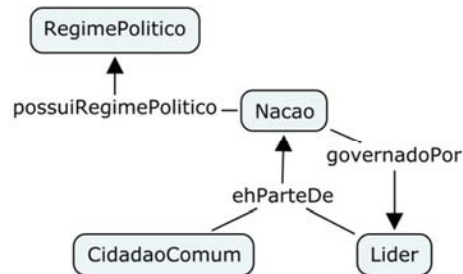



Figura 15 - Forma gráfica de uma conceitualização. (Fonte: a autora)

Apresentação formal

Linguagens para desenvolvimento de ontologias, como a OWL, são baseadas em formalismos lógicos. Para um programa que faz uso de inferências (*reasoner*), uma ontologia aparece como um conjunto de fórmulas lógicas que expressam os axiomas de uma teoria lógica (GRIMM et al, 2011). As linhas abaixo constituem um exemplo de conjunto de axiomas formalizados em lógica de primeira ordem[37], refletindo puramente o aspecto da representação do conhecimento, livre de aspectos sintáticos e ambiguidades presentes na linguagem natural.

Linguagem natural	Expressão lógica
Para toda nação existe um líder que a governa.	$\forall x$ Nacao(x) → $\exists y$(Lider(y) $\wedge$ governa(x, y))
Toda nação possui um regime político.	$\forall x$ Nacao (x) →$\exists y$(RegimePolitico(y) $\wedge$ possui(x, y))
Todo cidadão comum é um elemento da nação.	$\forall x$ CidadaoComum(x) → Nacao (x)
Todo líder é um elemento da nação.	$\forall x$ Lider (x) → Nacao (x)
Todo cidadão comum é submetido ao regime político da nação.	$\forall x$ CidadaoComum (x) → $\exists y$(RegimePolitico (y) $\wedge$ ehSubmetidoa(x, y))

Tabela 2 – Forma lógica de uma conceitualização (Fonte: a autora).

Apresentação para processamento por máquina

[37] Uma linguagem de primeira ordem encerra as seguintes categorias de símbolos lógicos:
(a) conectivos lógicos: ¬ (negação), $\wedge$ (conjunção/e), $\vee$ (disjunção/ou), → (implicação) e ↔ (equivalência).
(b) variáveis individuais: uma coleção infinita enumerável de variáveis individuais, que serão denotadas por x, y, z, w ...
(c) quantificadores: $\forall$ (para todo, qualquer que seja) e $\exists$ (existe);
(d) o símbolo de igualdade: =

Para ser processada por máquinas, uma especificação de ontologia precisa ser expressa em uma linguagem apropriada. O formato RDF e OWL embutido em marcações XML tem se apresentado como um formato adequado para a codificação de ontologias.

```
...
<owl:Class rdf:about ="#Nacao">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource ="#possuiRegimePolitico" />
            <owl:allValuesFrom rdf:resource ="#RegimePolitico" />
        </owl:Restriction>
    </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about ="#Cidadao">
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource ="#pertenceA" />
            <owl:someValuesFrom rdf:resource ="#NacaoBrasileira" />
        </owl:Restriction>
    </rdfs:subClassOf>
</owl:Class>
...
```

Figura 16 – Forma codificada em XML/ OWL de uma conceitualização (Fonte: a autora).

O trecho de código acima pode ser traduzido como: i) “Existe uma classe `Nação`, com a propriedade `possuiRegimePolitico`, cujos valores só podem vir da classe `RegimePolitico`”. ii) “Existe uma classe `Cidadão`, e alguns de seus indivíduos `pertenceA` `NaçãoBrasileira`”.

Para todos estes tipos de apresentação – gráfica, formal e para processamento - estão disponíveis ferramentas, a maioria *open sou*rce e gratuitas, que auxiliam na sua criação.

A seguir, abordaremos alguns aspectos que consideramos importantes para a contextualização do presente capítulo: o que motiva o uso de ontologias, do que ela se constitui, as linguagens e metodologias para construção, e as ontologias de fundamentação.

2.3.1 Por que ontologias?

O objetivo do uso de ontologias é promover um entendimento comum e compartilhado sobre um domínio. Uma série de razões descrevendo por que ontologias devem ser desenvolvidas é oferecida por alguns autores (NOY & McGUINNESS, 2001; MOREIRA et al, 2004; GRUBER, 1993) e nela estão incluídas: representação, reuso, compartilhamento,

aquisição e integração de conhecimento; extração de conceitos em textos de linguagem natural; comunicação entre sistemas, agentes e pessoas; recuperação de informação; especificação de software; e combinação de modelos de dados.

Nos últimos anos tem crescido o seu uso em sites de redes sociais e portais semânticos. Portais semânticos agregam funcionalidades semânticas de busca, publicação, colaboração e navegação taxonômica viabilizadas pelo uso de ontologias, além de se inserirem ao movimento global do *Linked Data*[38], de dados abertos e interconectados. Um projeto interessante neste sentido é o que está sendo desenvolvido pelo Centro de Pesquisa e Documentação de História Contemporânea do Brasil da Fundação Getulio Vargas em parceria com a Escola de Matemática Aplicada da mesma instituição (SOUZA et al, 2011). O projeto prevê o acesso unificado aos vários acervos históricos da casa, com navegação e busca pautada por conceitos (independente de mídias e de sistemas), além de integração a outros *datasources* na web (via *Linked Open Data*).

Uma lista de estudos de casos sobre portais semânticos apoiados por ontologias pode ser encontrada no site do grupo de trabalho sobre web semântica da W3C[39].

2.3.2 Do que se constitui uma ontologia

Para Gruber (1993), uma ontologia consiste em conceitos, definições e axiomas relativos a estes conceitos, dentro de um universo de discurso compartilhado. O autor identifica cinco tipos de construtos ontológicos: classes, que representam elementos de um domínio e que são definidas por um conjunto de atributos e seus possíveis valores; relações, que representam os tipos de associações entre elementos do domínio; funções, que são tipos especiais de relações que mapeiam um ou mais elementos do domínio para um elemento único[40]; axiomas, sentenças sobre o domínio que são sempre verdadeiras; e instâncias, que são os indivíduos da ontologia.

De acordo com Lassila e McGuinness (2001), uma ontologia deve possuir características: i) como obrigatórias: vocabulário controlado finito dos termos, que possa ser estendido; interpretação não-ambígua de classes e relacionamentos entre os termos do vocabulário; relacionamentos hierárquicos estritos entre classes e subclasses; ii) como típicas,

---

[38] *Sobre Linked Data* e *Linked Open Data*, ver a seção 5.1.3.
[39] http://www.w3.org/2001/sw/sweo/public/UseCases/
[40] Um exemplo de função é *ser-pais-biológicos*, onde um conceito homem e um conceito mulher estão relacionados a um conceito pessoa (GRUBER, 1993).

mas não obrigatórias: especificação de propriedades nas classes; inclusão de indivíduos (instâncias da ontologia); especificação de restrições; iii) como desejáveis, mas não obrigatórias nem típicas: especificação de classes disjuntas; especificação de relacionamentos lógicos arbitrários entre termos.

Na visão de Marcondes e Campos (2008), se considerarmos uma ontologia como um sistema de conceitos inter-relacionados, podemos dizer que estas contêm basicamente três componentes que são: um conjunto básico de conceitos e relações, uma forma de representação e um conjunto de assertivas.

Assim, destes e outros autores (GÓMEZ-PEREZ et al, 1996; NOY & McGUINESS, 2001; ALLEMANG & HENDLER, 2008) obtêm-se que ontologias são estruturas que funcionam basicamente através de:

- Classes;
- Propriedades;
- Restrições e regras – axiomas;
- Instâncias.

Classes – reúnem indivíduos que possuem algo em comum. Elas são restritas aos relacionamentos que estes participam, ou seja, é isso que irá definir uma classe em questão. A classificação hierárquica (superclasse-subclasse) estabelece relacionamentos de generalização e especialização (tem-tipo, é-um), onde passa a existir a presença de herança. Grande parte do poder das ontologias advém da capacidade de se raciocinar sobre os indivíduos que compõem o domínio, baseada nas propriedades conferidas pela classe ao qual pertencem.

Propriedades – são utilizadas para descrever as entidades do domínio. No domínio da história, para classes sobre eventos, possíveis propriedades seriam: `temDescrição`, `temDuracao`, `temEvidencia`, e `lideradoPor`. Na figura abaixo a propriedade `temIrmao` liga o indivíduo Getulio Vargas ao indivíduo Benjamim Vargas:

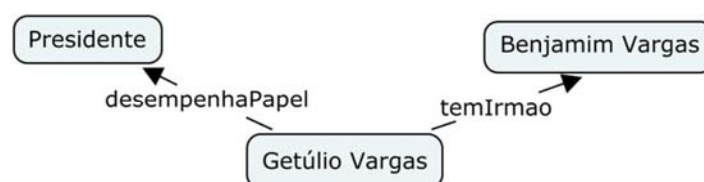



Figura 17 - Exemplo de representação de propriedades. (Fonte: a autora)

Propriedades possuem facetas para descrever seus valores, como o tipo, cardinalidade, domínio e escopo:

> O tipo de uma propriedade indica quais valores podem ser utilizados. Os tipos mais comuns são *texto* (qualquer sequência de caracteres), *número, lógico, pré-definido* (lista de opções pré-estabelecida) ou *instância* (relacionamento com outra classe).
>
> A cardinalidade de uma propriedade define a quantidade de valores que ela pode ter. Esta definição é similar à cardinalidade de modelos Entidade-Relacionamento de bancos de dados relacionais: uma propriedade com cardinalidade igual a 1 poderá conter apenas um único valor, enquanto que uma cardinalidade "n" implica na possibilidade de uma lista de diferentes valores. Por exemplo, a propriedade `temRepresentante` na classe `InstituiçãoPolitica` terá cardinalidade máxima de 513 para Câmara dos Deputados (uma vez que cada Estado pode eleger proporcionalmente até 70 deputados), e para Presidência da República, terá cardinalidade igual a 1.
>
> Domínio (*domain*) e escopo (*range*) especificam, respectivamente, o domínio de uma propriedade, ou seja, a classe de recursos que pode servir como sujeito da sentença e o alcance da propriedade, ou seja, a classe de recursos que pode servir como objeto da sentença. Por exemplo, na declaração "o suicídio aconteceu no Palácio do Catete", a relação "aconteceu em" tem como *domain* uma classe de evento e como *range* uma classe de lugar.

Relacionamentos são propriedades que estabelecem um tipo de interação entre as entidades de um domínio. Eles podem ser binários ou taxonômicos, temporais (implicam em precedência no tempo) ou topológicos (que expressam conexão espacial entre objetos). Nas relações binárias, são encontrados os seguintes tipos mais comuns: a relação classe-classe ('subclasse-de'), a relação instância-classe ('é-instância-de') e a relação instância-instância ('é-parte-de').

Restrições e **axiomas** – definem limites para os conceitos de um domínio, modelando verdades que sempre se cumprem na realidade representada pela ontologia. Servem como ponto inicial para dedução e inferências de outras verdades, além de validadores de sentenças. Um exemplo de axioma seria: 'todo evento na qual uma pessoa esteja envolvida deve ocorrer entre sua data de nascimento e data de falecimento'. Para Gruninger e Fox (1995) trata-se provavelmente da atividade mais difícil, porém necessária na definição de ontologias.

> Os axiomas na ontologia especificam a definição dos termos da ontologia e restringem a sua interpretação; são definidos em sentenças de primeira-ordem utilizando predicados [...] Simplesmente propor um conjunto de objetos sozinhos, ou propor um conjunto de termos em lógica de primeira-ordem, não constitui uma ontologia. Axiomas devem existir para definir a semântica, ou significado, destes termos (GRUNINGER e FOX, 1995)

Instâncias – são a materialização de uma classe. Enquanto classes definem coisas genéricas, instâncias de uma determinada classe representam indivíduos específicos daquela classe. Um exemplo de instância da classe `InstituiçãoPolítica` poderia ser `Câmara dos Deputados`: através da classe a ontologia sabe que ela é uma instituição política e através da instância sabe qual instituição especificamente se trata, diferenciando-a das outras instituições.

2.3.3 Linguagens e esquemas

Revendo o conjunto de especificações para a web semântica (seção 5.1.4), tem-se que da sua arquitetura fazem parte as seguintes tecnologias para a representação dos recursos (BERNERS-LEE, 2001): camada de sintaxe (XML/XMLSchema), camada de dados (RDF/RDF Schema), camada de ontologias (OWL), camada de lógica e camada de inferência.

Vimos que por si só os metadados são insuficientes para a construção da web semântica. Para que sejam codificados e manipulados por agentes, é necessário ir além da análise sintática dos dados e criar uma estrutura semântica que descreva formalmente os significados subjacentes aos conteúdos que representam. As linguagens de representação do conhecimento tomam para si esta tarefa.

Formalização dos significados em RDF

Significados podem ser expressos através de RDF – *Resource Description Framework*. A ideia do RDF é descrever dados por meio de um esquema de “triplas” de recurso-propriedade-valor – ou sujeito-predicado-objeto –, expressas em sentenças escritas em XML (SOUZA e ALVARENGA, 2004). Em RDF, um documento pode fazer asserções do tipo: alguma coisa (pessoa, objeto, acontecimento) tem propriedade (participou de, tem relação) com certo valor (outra pessoa, outro objeto).

```
<rdf:RDF>
  <rdf:Statement>
    <rdf:subject   rdf:resource="Getulio Vargas" />
    <rdf:predicate rdf:resource="temIrmão" />
    <rdf:object    rdf:resource="Benjamim Vargas" />
  </rdf:Statement>
</rdf:RDF>
```

Figura 18 – Exemplo de asserções em RDF

Cada um dos recursos com suas propriedades é identificado por uma URI exatamente como vimos no XML, e isso permite definir a qualquer momento novos recursos ou propriedades bastando para isso definir outras URIs (BERNERS-LEE, HENDLER e LASSILA, 2001).

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    xmlns:hist="http://pesquisa.dominio.br/historia">
  <foaf:Person rdf:about="http://pesquisa.dominio.br/historia/#getuliovargas">
    <foaf:name=xml:lang="pt">Getúlio Dornelles Vargas</foaf:name>
    <hist:temIrmao rdf:resource="http://pesquisa.dominio.br/historia/#benjamimvargas" />
  </foaf:Person>
</rdf:RDF>
```

Figura 19 - Exemplo de RDF embutido em XML

O exemplo acima faz referência a três diferentes *namespaces*: os dos padrões compartilhados RDF[41] e FOAF[42] e do fictício criado como exemplo, o HIST. Uma vez especificado um *namespace*, já é possível utilizar seus elementos de forma não ambígua, bastando referenciá-los corretamente: `foaf:Person` assevera que o sujeito da sentença é uma pessoa; `foaf:name`, que essa pessoa possui um nome; `hist:temIrmão`, que ela tem um irmão; e `rdf:resource`, que esse irmão é identificado pela URI tal. Assim, cada *namespace* representa um vocabulário compartilhado de conceitos e propriedades bem definidos, alguns bastante genéricos e abrangentes, outros de domínios específicos.

O RDF Schema (RDFS) estende o RDF com novos elementos, acrescentando ao modelo a possibilidade de obter uma descrição mais rica e estruturada com o uso de classes, hierarquias e a noção de tipos de dados (`rdfs:class`, `rdfs:subclass`, `rdfs:DataType`). Possui ainda recursos para determinar as propriedades adequadas a uma classe

[41] http://www.w3.org/RDF/
[42] http://xmlns.com/foaf/spec/

(rdfs:domain)e valores adequados a uma propriedade (rdfs:range) (DOMINGUE, FENSEL & HENDLER, 2011; ALMEIDA & SOUZA, 2011).

Para ilustrar, podemos definir que Senado Federal é uma instituição política que integra o Congresso Nacional, e da qual participa o indivíduo Francisco Dornelles, representando o estado do Rio de Janeiro. A representação de um grafo RDF contendo as relações entre estes conceitos seria algo como:

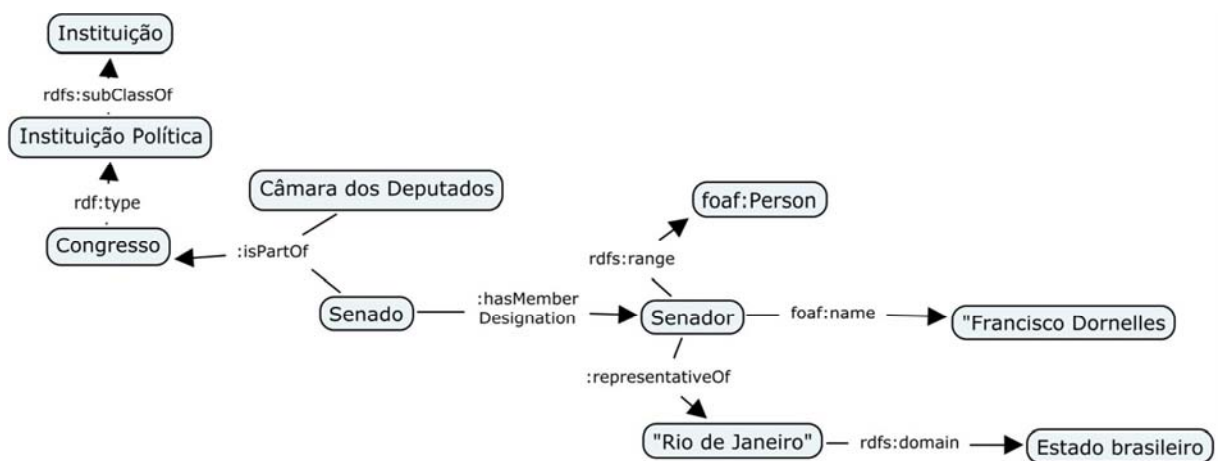



Figura 20 – Exemplo de asserções RDF na forma de grafo (baseado em ALLEMANG & HENDLER, 2008)

Assim, o RDF e o RDFS fornecem um conjunto de primitivas para a modelagem de ontologias simples. No entanto, conforme Allemang & Hendler (2008) apontam, não é costume adotá-los isoladamente por existirem lacunas de expressividade de seus construtos. A crítica recai principalmente na inexistência de relações mais complexas entre propriedades (como a relação simétrica, inversa e transitiva), entre classes (como disjunção e equivalência) e também de restrições de valor e cardinalidade para as propriedades.

A OWL (*Ontology Web Language)* visa preencher estas lacunas, e desde 2004 é uma linguagem de representação recomendada pela W3C para o desenvolvimento de ontologias.

Construção de ontologias com OWL

A OWL é originada a partir da junção de duas outras linguagens, a DAML e OIL[43], e fornece uma riqueza maior de construtos do que a provida pelo RDF e RDFS. Ela é apresentada em três versões, que variam de acordo com a expressividade da linguagem: OWL Lite, OWL DL e OWL Full.

[43] Para mais informações ver: http://en.wikipedia.org/wiki/DAML+OIL

- OWL Lite - voltada para projetos que demandem uma classificação hierárquica simples, possui menor complexidade formal. A sua adoção permite, inclusive, a migração rápida de tesauros e outras taxonomias para o formato de ontologias. Embora suporte restrições de cardinalidade, elas se restringem a valores de 0 ou 1.

- OWL DL – esta versão é baseada no formalismo da lógica descritiva (DL), fornecendo construtos desejáveis em sistemas que necessitem de um maior nível de detalhamento e restrições. No entanto, ela exige a separação de tipos (classe, propriedade e indivíduo), ao contrário da OWL Full. Permite cardinalidades mínima, máxima e intervalos.

- OWL Full – garante o máximo de expressividade e a liberdade sintática do RDF. Uma classe, por exemplo, pode ser tratada simultaneamente como uma coleção de indivíduos, ou simplesmente como um indivíduo. Esta versão permite aumentar o significado de um vocabulário (RDF ou OWL), porém sem nenhuma garantia computacional (completude, decidibilidade, raciocínio).

Segundo o guia publicado por Smith et al (2004) e disponível no site da W3C[44], os principais construtos de uma ontologia em OWL são basicamente as classes, as propriedades e as restrições:

1. <u>Classes:</u> no mundo OWL existe uma classe superior da qual todos os indivíduos são membros, é a classe `owl:Thing`. Todas as outras que constituirão as taxonomias do domínio são implicitamente subclasses dela.

```
<owl:Class rdf:ID="InstituicaoPolitica">
  <rdfs:subClassOf rdf:resource="#Instituicao" />
  ...
</owl:Class>
```

Figura 21 - Exemplo de definição de classes em OWL

Quando duas classes não podem ter subclasses ou instâncias em comum, diz-se que elas são disjuntas: a classe Homem é disjunta da classe Mulher (`owl:disjointWith`). Quando duas classes são conhecidas por sempre conterem os mesmos membros, diz-se que elas são equivalentes: País é equivalente à Nação; ou

[44] Disponível em: http://www.w3.org/TR/owl-guide/

Regime Militar é equivalente à Ditadura Militar e equivalente à Governo Militar (`owl:equivalentClass`).

2. Propriedades OWL representam relacionamentos. Existem dois tipos principais de propriedades: de objetos (`owl:objectProperty`) e de dados (`owl:datatypeProperty`). Os primeiros ligam um indivíduo/instância a outro, e os segundos descrevem relacionamentos entre indivíduos e valores de dados (HORRIDGE et al, 2007). A figura 17 ilustra esta diferença:

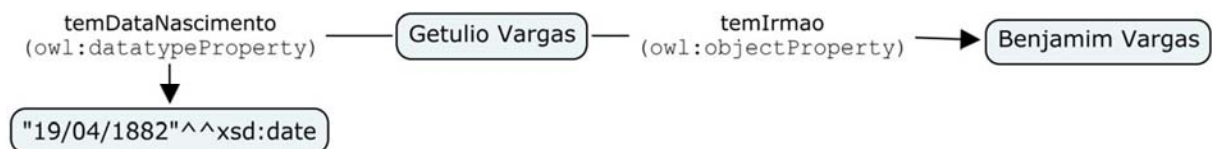



Figura 22 - owl:objectProperty e owl:datatypeProperty

Tipos de propriedades de objetos (HORRIDGE et al, 2007; ALLEMANG & HENDLER, 2008):

a) Inversa (`owl:inverseOf`): alterna sujeito e objeto da tripla. Por exemplo, se Getulio Vargas escreveu Carta-Testamento, então Carta-Testamento foi escrita por Getulio Vargas. Na notação N3[45]:

```
    :GetulioVargas :escreveu :Carta-Testamento
:escreveu owl:inverseOf :foiEscritoPor
→ :Carta-Testamento :foiEscritoPor :GetulioVargas
```

b) Transitiva (`owl:transitivProperty`): cadeias de relacionamento são colapsadas em um único relacionamento. Por exemplo, se a propriedade ter parentesco é transitiva; e Getulio tem parentesco com Lutero; e tem parentesco com Alzira; então Lutero necessariamente tem parentesco com Alzira.

```
:temParentesco rdf:type owl:transitiveProperty
:GetulioVargas :temParentesco :LuteroVargas
:GetulioVargas :temParentesco :AlziraVargas
→ :LuteroVargas :temParentesco :AlziraVargas
```

c) Simétrica (`owl:symmetricProperty`): a propriedade que é o seu próprio inverso. Assim, o inverso de ter um cônjuge é o próprio ter um cônjuge:

```
:temConjuge rdf:type owl:symmetricProperty
```

[45] *Notation 3 RDF*, ou notação N3, possibilita expressar textualmente e de forma simples as triplas sujeito-predicado-objeto das sentenças RDF. Mais informações em: http://www.w3.org/2000/10/swap/Primer.

d) Anti-simétrica (`owl:antisymmetricProperty`): se um indivíduo é ligado a outro por uma propriedade, por exemplo, `éFilhoDe`, este outro não pode ser ligado àquele pela mesma propriedade.

e) Funcional (`owl:functionalProperty`): apenas um valor é permitido como objeto. Por exemplo, um indivíduo só pode ter *uma* mãe biológica:

```
:temMaeBiologica rdf:type owl:functionalProperty
:GetulioVargas :temMaeBiologica :CandidaDornelles
:GetulioVargas :temMaeBiologica :CandidaVargas
→ :CandidaDornelles owl:sameAs :CandidaVargas
```

f) Funcional Inversa (`owl:functionalProperty`): apenas um valor é permitido como sujeito. Assim, `:éMãeBiológica` é a propriedade inversa de `:temMaeBiologica`, porque esta última é funcional.

3. Restrições, ou axiomas: de acordo com Allemang & Hendler (2008), uma restrição `owl:Restriction` é um tipo especial de classe, que contém indivíduos como qualquer outra. Os membros dessa classe precisam satisfazer a condição especificada pelo tipo de restrição (`owl:someValuesFrom`, `owl:allValuesFrom` e `owl:hasValue`) aplicada à propriedade (`owl:onProperty`).

a) `owl:someValuesFrom` – é usado para produzir uma restrição na forma 'todos os indivíduos no qual pelo menos um (*at least one/some*) valor da propriedade P vem da classe C'. No exemplo abaixo, a restrição formulada gera a classe dos indivíduos que possuem pelo menos um relacionamento `ocupouCargo` com os membros de `CargoDeOrgaoGovernamental`:

```
[a owl:Restriction;
   owl:onProperty :ocupouCargo;
   owl:someValuesFrom :CargoDeOrgaoGovernamental]
```

b) `owl:allValuesFrom` – é usado para produzir uma restrição na forma 'os indivíduos no qual todos (*all/only*) os valores da propriedade P vem da classe C. Por exemplo, a classe de indivíduos que possui apenas relacionamentos `ocupouCargo` com os membros de `CargoDoLegislativo`:

```
[a owl:Restriction;
   owl:onProperty :ocupouCargo;
   owl:allValuesFrom :CargoDoLegislativo]
```

c) `owl:hasValue` – é usado para produzir uma restrição na forma ‘os indivíduos que possuem o valor x para a propriedade P’.

```
[a owl:Restriction;
   owl:onProperty :ocupouCargo;
   owl:hasValue :Senador]
```

A seguir, um resumo dos construtos OWL, presentes nas versões Lite, DL e Full, tendo em mente que: a) toda ontologia OWL Lite válida é uma ontologia OWL DL válida; b) toda ontologia OWL DL válida é uma ontologia OWL Full válida; c) porém essas relações não são inversamente verdadeiras.

Construtos OWL Lite, OWL DL, OWL Full

RDF Schema (RDFS)

- `Class`
- `rdfs:subClassOf`
- `rdf:Property`
- `rdfs:subPropertyOf`
- `rdfs:domain`
- `rdfs:range`
- `Individual`

Igualdade e diferença

- `equivalentClass`
- `equivalentProperty`
- `sameAs`
- `differentFrom`
- `AllDifferent`
- `distinctMembers`

Características de Propriedade

- `ObjectProperty`
- `DatatypeProperty`
- `inverseOf`
- `TransitiveProperty`
- `SymmetricProperty`
- `FunctionalProperty`
- `InverseFunctionalProperty`

Restrições de Propriedade

- `Restriction`
- `onProperty`
- `allValuesFrom`
- `someValuesFrom`
- `hasValue`

Restrições de Cardinalidade

- exatamente 0 ou 1
- `minCardinality` (ao menos)
- `maxCardinality` (ao máximo)
- `cardinality` (intervalo)

Construtos de Classes

- `intersectionOf`
- `oneOf`
- `unionOf`
- `complementOf`

Restrições sobre Classes

- `disjointWith`
- Classes complexas: combinações mais ricas envolvendo construtos de classes, restrições de propriedades e expressões booleanas
- Classes podem ser tratadas como instâncias

Tabela 3 - Resumo dos construtos em OWL Lite, DL e Full
(Fonte: http://www.w3.org/TR/2004/REC-owl-features-20040210/#s2.1)

2.3.4 Metodologias de construção de ontologias

Existem inúmeras metodologias que contemplam as atividades envolvidas no processo de construção de ontologias. Não há como qualificá-las, pois em geral são concebidas para propósitos particulares e em circunstâncias distintas. Como apontam Noy e McGuiness (2001, p. 4), “não existe um único modo correto para modelar um domínio – existem alternativas viáveis. A melhor solução quase sempre depende da aplicação que se tem em mente”.

Algumas metodologias, como a apresentada por Gruninger e Fox[46] (1995), a Methontology[47] de Gómez-Pérez et al (1996, 1997), o método 101[48] de Noy e McGuinness (2001) e a Ordnance Survey[49] (KOVACS et al, 2006) foram estudadas com mais detalhes. Clareza na sistematização das etapas, existência de documentação e citações na área foram os critérios para esta seleção, que resultou na separação de alguns princípios comuns que parecem ser representativos do processo. Em Silva (2008) são tecidas considerações sobre a análise comparativa entre estas e outras metodologias tendo por base um arcabouço teórico-conceitual recuperado das ciências da informação e da computação, resultando em uma proposta-síntese de metodologia consistente.

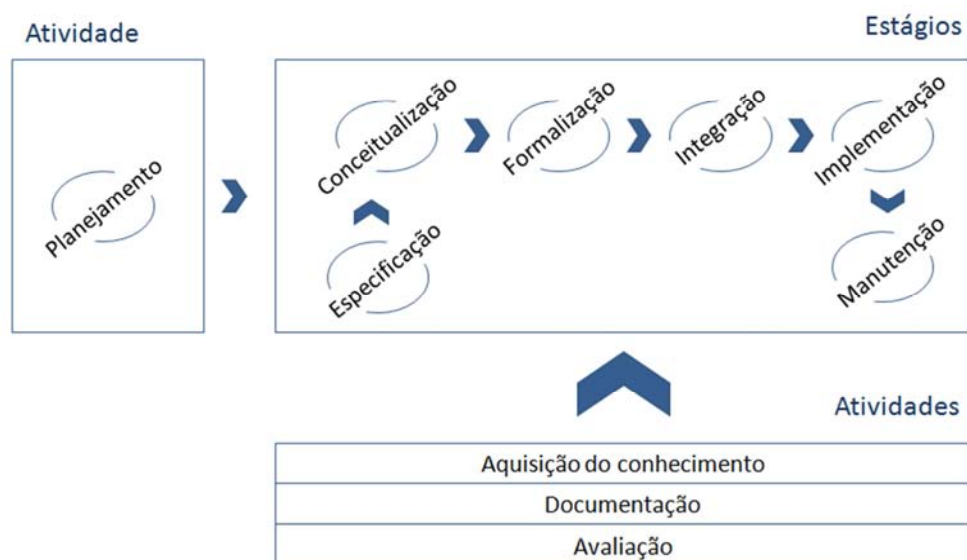



Figura 23 - Etapas de desenvolvimento de ontologias (Adaptado de Gómez-Pérez et al, 1997)

[46] Os autores conceberam esta metodologia no contexto de um projeto voltado para representações do comportamento empresarial, o projeto TOVE (Toronto Virtual Enterprise).
[47] A Methontology é fruto da experiência dos autores na construção de uma ontologia no domínio da Química.
[48] O Método 101 é apresentado no contexto de uma ontologia de vinhos e alimentos.
[49] O guia de referência do Ordnance Survey é baseado na construção de uma ontologia de Hidrologia.

Cada autor se detém no que considera mais importante para a qualidade da sua ontologia, porém todos contemplam, em menor ou maior profundidade, etapas de especificação, aquisição do conhecimento, conceitualização, formalização, integração, implementação e avaliação da ontologia. Muitos dos princípios reunidos são provenientes de campos mais maduros, como a engenharia de software e engenharia do conhecimento (GÓMEZ-PÉREZ, 1999).

Fase de especificação

O ponto de partida para a construção de uma ontologia reside, em primeiro lugar, na identificação clara do escopo, propósitos e usos esperados para ela. Qual domínio a ontologia cobrirá? Que propósito visa atender? A quem se destina e que questões de competência ela pretende responder?

Com base na elaboração de cenários de motivação (Gruninger & Fox, 1995), as questões de competência justificam e ajudam na escolha dos conceitos básicos da ontologia, podendo servir posteriormente para avaliar tanto os compromissos ontológicos[50] estabelecidos quanto sua própria expressividade. Estas questões podem ser apenas um roteiro e não precisam ser exaustivas.

A Methontology prevê ainda para esta fase a especificação do nível de formalidade esperado, as características da terminologia e granularidade.

Aquisição do conhecimento

Para Gómez-Pérez et al (1996, 1997), a aquisição de conhecimento é uma atividade independente do restante do processo, mas que acontece concomitante à fase de especificação e avança em menor grau por todo o ciclo de desenvolvimento. Nesta fase os termos principais do domínio são identificados e reunidos. Fontes de conhecimento podem ser os especialistas, livros, documentos, tabelas de dados e outras ontologias, e técnicas como *brainstormings*, entrevistas e análise de textos.

[50] O compromisso ontológico é um acordo firmado por uma comunidade sobre o significado que ela aquiesce em reconhecer acerca dos conceitos de uma ontologia, tanto do ponto de vista da sua compreensão pelo homem quanto do seu tratamento pela máquina. Dessa forma é possível estabelecer uma compreensão mais próxima do sentido original pretendido naquilo que está sendo representado (NODINE E FOWLER , 2002, apud CAMPOS, 2011).

Na área de processamento de linguagem natural, técnicas de aprendizado de máquina para a extração automática de entidades nomeadas têm sido desenvolvidas com o intuito de, entre outras coisas, auxiliar na aquisição de conhecimento para ontologias e atribuir anotações semânticas aos textos (PATTUELLI, 2007; BONTCHEVA & CUNNINGHAM, 2011). Iniciativas como o *Automatic Content Extraction Evaluation*[51] (ACE) têm orientado o desenvolvimento de ferramentas para aplicações reais e de pesquisa, cujas principais atribuições são:

- Reconhecimento de entidades nomeadas, que consiste na identificação e classificação de diferentes tipos de nomes presentes no texto;
- Resolução de co-referências, que é a tarefa de decidir se duas expressões lingüísticas se referem ou não à mesma entidade no discurso;
- Extração de relacionamentos, que identifica as relações entre entidades no texto.

Conceitualização

Em seguida é hora de organizar esse conhecimento separando os conceitos, instâncias, verbos e propriedades, para, a partir deles construírem-se representações intermediárias para o domínio. A Methontology sugere a elaboração de uma árvore de classificação com a disposição dos conceitos em suas relações taxonômicas hierárquicas e partitivas. Feito isso, parte-se para a criação de um dicionário de dados contendo a descrição dos conceitos, seus significados, exemplos de instâncias e propriedades. Separam-se os conceitos – termos que existem de forma independente – das propriedades – termos que dependem de algum conceito para existir (NOY & McGUINNESS, 2001). Também faz parte deste processo descrever os axiomas, fórmulas e regras que serão usados no domínio.

---

[51] http://www.ldc.upenn.edu/Projects/ACE/

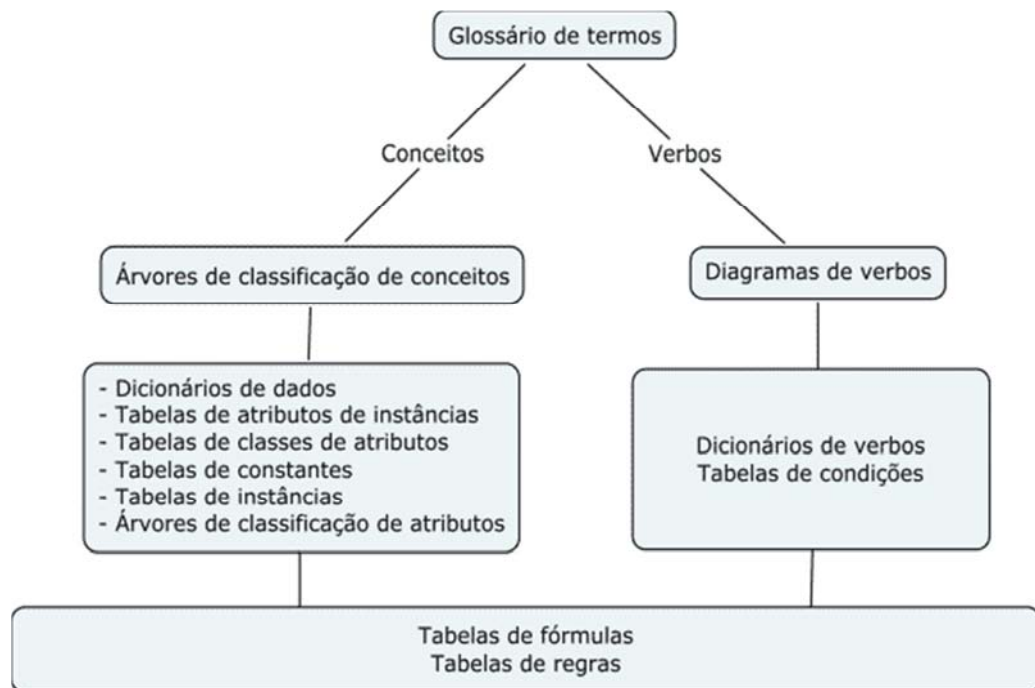



Figura 24 - Representações conceituais intermediárias na Methontology
(Fonte: Gómez-Perez et al, 1997 apud Silva, 2008)

Esta etapa produz modelos de representações que são independentes de linguagem, e, portanto, inteligíveis para o especialista do domínio.

As metodologias apontam as formas possíveis para a concepção das classes e subclasses. A abordagem denominada *top-down* privilegia as definições mais genéricas para depois seguir especializando, já a abordagem *bottom-up* inicia pelas definições mais específicas a fim de agrupá-las em categorias genéricas. Uma combinação das duas seria identificar primeiramente os conceitos que mais se destacam e, em seguida, generalizar ou especializar. Como os conceitos *in the middle* tendem a ser os mais descritivos do domínio, esta costuma ser a abordagem mais intuitiva para a maioria dos desenvolvedores.

Na Ordnance Survey, o conhecimento capturado é estruturado em tabelas que descrevem os conceitos (dicionário de termos), os tipos de relacionamentos presentes (herança, mereologia, contenção, etc) e as regras de relacionamento (inversa, transitiva, simétrica, etc). A etapa de conceitualização, quando bem implementada, minimiza as chances de apresentação de inconsistências e ambiguidades conceituais.

Fase de formalização da ontologia

Nesta etapa são definidos formalmente os termos da ontologia e as suas restrições de interpretação correspondentes. A metodologia de Gruninger e Fox faz uso da lógica de

primeira ordem, ou o equivalente em KIF[52] para descrever os objetos e propriedades da ontologia. A linguagem formal proverá a terminologia necessária para reafirmar as questões de competência e expressará as definições e restrições nos axiomas. Os objetos são representados por constantes e variáveis, as propriedades de atributo são definidas por predicados unários[53] e as relações por n-ários predicados.

Nesta linguagem, em contraste com a linguagem natural, usam-se símbolos não ambíguos e formulações exatas que permitem que deduções sejam testadas com maior precisão. O tratamento lógico/teórico de qualquer domínio consiste em propor sentenças sobre os objetos neste domínio (sentenças atribuindo certas propriedades e relações aos objetos em questão) e em estabelecer regras de acordo com as quais outras sentenças possam ser derivadas (GUIZZARDI, 2000).

A proposição dos termos em lógica não constituem já uma ontologia, antes, ela provê a semântica do domínio e restringe sua interpretação. A definição dos axiomas é um processo iterativo guiado pelas questões de competência formais. Para muitos autores é a etapa mais difícil na modelagem de ontologias.

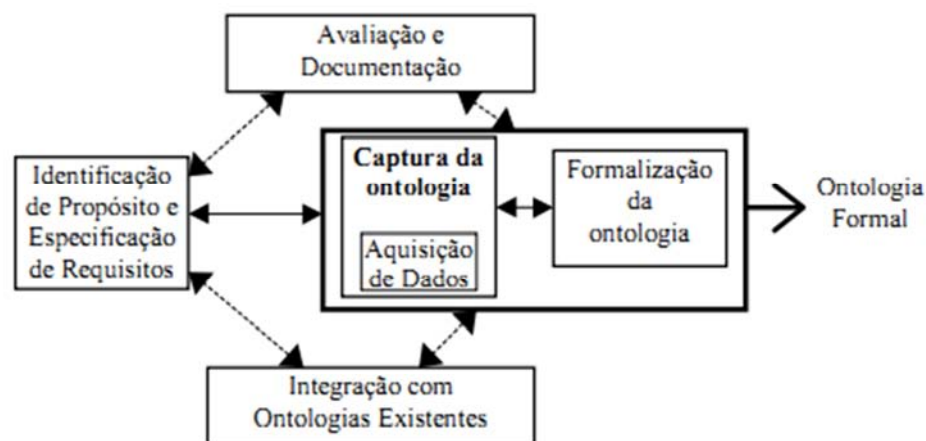



Figura 25 – Ciclo interativo das etapas de construção de ontologias (Extraído de Guizzardi, 2000)

[52] *Knowledge Interchange Format* é uma linguagem para o intercâmbio de conhecimento entre diferentes programas de computador. Possui semântica declarativa, isto é, o significado das expressões pode ser entendido sem necessidade de um interpretador. (Fonte: Wikipedia)

[53] Exemplos de predicados unários seriam: '... foge', '... é branco', '... é belo'; predicados binários: '... ama ...', '... é mais pesado que ...'; predicados trenários: '... fica entre ... e ...'.

Integração com ontologias existentes

O reuso de ontologias é recomendável sempre que possível. A integração e alinhamento com outras ontologias de topo ou domínio podem significar redução no esforço de criação e aumento do nível de interoperabilidade. A Methontology sugere a criação de um documento que contenha para cada definição emprestada, o nome da ontologia de origem, o nome do termo no original e seus argumentos.

Implementação

Segundo Gómez-Pérez et al (1996, 1997), esta fase diz respeito a codificação da ontologia em uma linguagem de representação formal como LOOM, Prolog, Ontolingua, C++ ou outra. Um ambiente adequado de desenvolvimento de ontologia, segundo os autores, deve prover no mínimo: i) analisador léxico e sintático para garantir a ausência de erros nesta área; ii) tradutores para garantir a portabilidade das definições em outras línguas; iii) editor para incluir, remover ou modificar definições; iv) navegador para examinar a biblioteca de ontologias e suas definições; v) buscador para pesquisar as definições mais apropriadas; vi) módulo de avaliação para detectar incompletude, inconsistências e conhecimento redundante; vii) mantenedor automático para gerenciar a inclusão, remoção ou modificação de definições existentes.

O Método 101 apresenta a implementação de sua ontologia de estudo de caso no editor Protegé[54], que possui suporte para as linguagens de representação XML, RDF e OWL.

Avaliação

Gruninger e Fox (1995) apontam a importância da verificação de completude, ou seja, do atendimento pleno das questões de competência como fator de avaliação da ontologia. Isto pode ser feito através de teoremas que fornecem as condições necessárias para esta verificação.

Já Gruber (1993), acredita que uma ontologia deve contemplar alguns critérios fundamentais para ser considerada de qualidade, quais sejam: clareza, coerência, extensibilidade, codificação mínima e compromissos ontológicos mínimos. O cotejamento destes critérios fornece elementos para avaliação.

---

[54] http://protege.stanford.edu/

Para Gómez-Pérez et al (1996, 1997) esta é uma atividade a ser realizada durante cada fase e entre fases do ciclo de vida da ontologia, para que possam ser garantidas a exatidão e correspondência do resultado frente à proposta inicial do desenvolvimento.

Documentação

A Methontology estabelece a documentação como uma atividade a ser realizada também durante todo o processo de desenvolvimento da ontologia. O resultado é a reunião de um conjunto de documentos contendo informações sobre: i) especificação de requisitos; ii) aquisição do conhecimento; iii) modelo conceitual e representações intermediárias para descrição do domínio; iv) formalização; v) integração; vi) implementação; vii) avaliação.

O editor de ontologias Protegé fornece campos para entrada de descrições de cada elemento inserido na ontologia (classes, propriedades, restrições) em linguagem natural. Assim, se o usuário preencher corretamente estes campos, a documentação é gerada automaticamente.

Algumas considerações

É importante ressaltar que, além das metodologias mencionadas, as várias experiências reais de construção de ontologias visitadas foram igualmente importantes para a constituição de nosso próprio método de trabalho. A maioria delas produziu farta documentação na forma de artigos e guias, que possibilitaram o cotejamento entre a orientação adotada e os resultados alcançados, tornando o processo de desenvolvimento intuitivamente mais claro. Estas experiências serão detalhadas no capítulo 4.

2.3.5 Ontologias de Fundamentação

Ontologias de fundamentação, ou de topo, são sistemas de categorias filosoficamente bem fundamentados e independentes de domínio (GUIZZARDI, FALBO & GUIZZARDI, 2008). Em uma abordagem *top-down* podem fornecer um ponto de partida para a modelagem de um domínio.

Porém, mais do que apenas descrever conceitos gerais, elas provêem axiomatizações ricas que explicitam e fundamentam acordos ontológicos. Isto se torna especialmente útil em projetos onde são previstos adoção e reuso de múltiplas ontologias indepedentes, onde os termos devem ser definidos de forma bastante precisa a fim de evitar ambiguidades terminológicas e conceituais durante a integração destas ontologias.

Face à abordagem pretendida para o domínio chega-se a diferentes ontologias de fundamentação – mesmo que algumas carreguem princípios semelhantes. Elas fornecem caminhos para escolhas do tipo (MASOLO et al, 2003; SMITH, 2004; GUIZZARDI, 2005; OBERLE et al, 2006): de que forma as entidades serão dependentes da noção espaço-tempo? Atributos serão tratados como qualidades, tropes ou uma espécie de continuantes? Propriedades essenciais conferirão à ontologia uma abordagem multiplicativa de suas entidades? Como serão modelados conceitos de modo ou de coisas possíveis, mas não reais?

Uma ontologia de fundamentação é considerada, portanto, uma teoria axiomática sobre categorias de topo (ou primitivas) independentes de domínio que encontramos no mundo, tais como objeto, atributo, evento e conexões espacial e temporal (OBERLE et al, 2006; PROBST, 2006). Por carregarem pressupostos de natureza muito abstrata, sua compreensão não é tão simples assim, sendo necessário algum conhecimento de filosofia.

De fato, alguns autores, como Keet (2011), apontam que ontologias de fundamentação são deveras complexas para o desenvolvimento de ontologias simples (*lighweight*) de domínio, e que se leva muito tempo para entendê-las em detalhes. E ainda, que o seu uso não assegura a qualidade ou eficiência na construção de ontologias[55]. Mas a verdade é que atualmente trata-se de prática bastante adotada, conforme a literatura relata (OBERLE et al, 2006; GUIZZARDI et al, 2008; SCHERP et al, 2009; SHAW et al, 2009; VAN HAGE et al, 2009). Os principais argumentos residem, sobretudo, na promessa de redução do esforço de ‘pensar’ a criação de categorias gerais e propriedades básicas, e na interoperabilidade ao permitir aderência a outras ontologias de mesma base.

Existem várias ontologias de fundamentação disponíveis na Internet e gratuitas, como a DOLCE *Descriptive Ontology for Linguistic and Cognitive Engineering* (GANGEMI et al., 2002; MASOLO et al, 2003), a BFO *Basic Formal Ontology* (MASOLO et al, 2003), a SUMO *Suggested Upper Merged Ontology* (NILES, PEASE, 2001; OBERLE et al, 2006) e a UFO *Unified Foundational Ontology* (GUIZZARDI, WAGNER, 2010). Argumentos para a escolha de uma ou outra são discutidos na literatura da área (OBERLE et al, 2006; SCHERP et al, 2009; VAN HAGE et al, 2011).

---

[55] Esta autora realizou um teste com um grupo de desenvolvedores, cuja proposta era criar uma ontologia. Alguns partiram do zero e outros de ontologias de fundamentação. O resultado mostrou que este último modelo nem sempre é a melhor escolha em termos de qualidade e tempo de desenvolvimento. Mais em (KEET, 2011).

2.3.6 Ontologias de Domínio

As ontologias do domínio capturam o conhecimento válido para um tipo particular de domínio. São concebidas a partir do conhecimento depreendido dos especialistas do domínio em questão, sendo necessária uma profunda análise que revele seus conceitos, atributos e relações mais relevantes, além das restrições e regras que os governam (STUDER, BENJAMINS & FENSEL, 1998).

Os conceitos nas ontologias de domínio são frequentemente definidas como especializações dos conceitos gerais que existem nas ontologias de referência. Por exemplo, para o domínio da História, podemos incluir conceitos tais como ditadura e regime militar, que são especializações de uma ontologia genérica. A fronteira entre os dois níveis é bastante vaga, mas a distinção é intuitivamente significativa e útil para a construção de bibliotecas (VAN HEIJST, SCHREIBER & WIELINGA, 1997).

Uma grande quantidade de ontologias de domínio tem sido desenvolvida para domínios como medicina, direito, engenharia, modelagem organizacional e química (GUIZZARDI, 2005). Sites como *Some Ongoing KBS/Ontology Projects and Groups*[56] da Universidade do Texas, trazem trabalhos e projetos que utilizam ou desenvolvem ontologias de domínio.

Integração de ontologias

A figura 26, extraída de (SCHERP, 2012), mostra como as ontologias se integram em um modelo de representação:

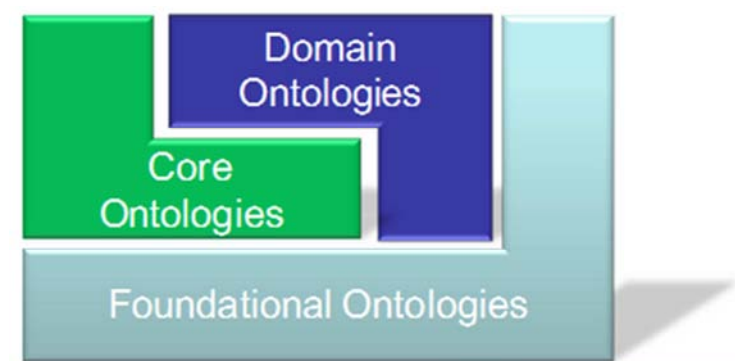



Figura 26 - Pilha de ontologia (SCHERP, 2012)

Como podemos observar, as ontologias de fundamentação ou de topo (*foundational ontologies*) são transversais e alcançam múltiplas ontologias de núcleo e de domínio. As ontologias de núcleo ou de referência (*core ontologies*) também atuam, de certa forma, como

[56] http://www.cs.utexas.edu/users/mfkb/related.html

as ontologias de topo, porém elas definem um conjunto de entidades selecionadas para determinados modelos conceituais, como: estado, evento, processo, ação, tempo, lugar. Já as ontologias de domínio são obtidas da especialização das anteriores, fornecendo um vocabulário específico com conceitos próprios. Como exemplo, temos ontologias que versam sobre futebol, sobre pesca, sobre a história do Brasil, sobre anatomia, etc.

Alinhamento entre ontologias de topo e de domínio

Segundo Probst (2006) as três situações possíveis no processo de alinhamento de ontologias podem ser resumidas da seguinte forma:

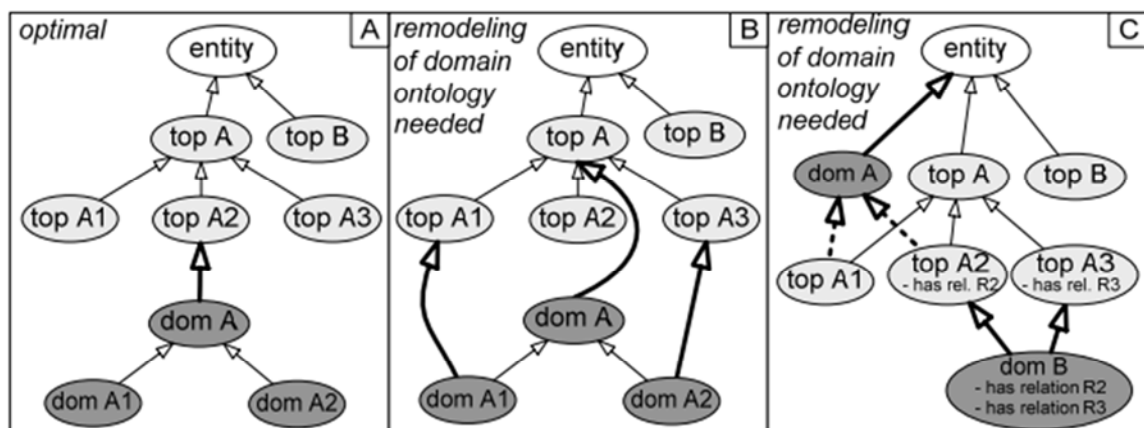



Figura 27 - as possíveis situações para o alinhamento entre ontologias de topo e de domínio.
Fonte: Probst (2006)

A. Os conceitos mais gerais da ontologia de domínio são alinhados aos conceitos mais específicos da ontologia de fundamentação; seria a situação ótima, e deveria ser o resultado de um processo de alinhamento;

B. Relações taxonômicas da ontologia de domínio se sobrepõem às relações taxonômicas da ontologia de fundamentação. Neste caso, a ontologia de domínio proverá conceitos que são mais gerais que os conceitos mais específicos da ontologia de fundamentação;

C. Um conceito da ontologia de domínio pode atuar como superclasse de vários conceitos da ontologia de fundamentação (conceito “dom A” da figura). Como esta generalização não é considerada na ontologia de fundamentação, isto pode indicar uma super generalização na ontologia de domínio.

Vimos nesta seção, portanto, que as ontologias como instrumento de representação são estruturas formais concebidas a partir de um rico arsenal teórico metodológico, subjacentes às escolhas que se fazem sobre o domínio que se quer representar.

2.4 O domínio da História

> “Estudar história significa entregar-se ao caos, conservando a crença na ordem e no sentido”. (Herman Hesse, *O Jogo das Contas de Vidro*)

O historiador medievalista francês Marc Bloch (1886-1944) menciona dois fenômenos que para ele estão no cerne da História: a duração, matéria concreta do tempo, e a aventura, forma individual e coletiva da vida dos homens, arrastados ao mesmo tempo pelos sistemas que os superam e confrontados a um acaso no qual se exprime com frequência a mobilidade da história (BLOCH, 2001). Assim, o grande e principal objetivo da ciência histórica é trazer à luz a versão mais plausível possível de períodos ou acontecimentos relevantes da humanidade, seja na sua dimensão macro ou delimitados em pequenos recortes temáticos. E para Bloch, isso é algo a ser realizado com o cuidado e a sensibilidade de um maestro a compor sua obra: “É que o espetáculo das atividades humanas, que forma seu objeto específico, é, mais que qualquer outro, feito para seduzir a imaginação dos homens” (BLOCH, 2001, p. 44). Antes do desejo de conhecimento, há o estímulo pelo ‘simples gosto’; distrair é uma das virtudes da História. E assim segue o autor: “Evitemos retirar de nossa ciência sua parte de poesia” (BLOCH, 2001, p. 44). A História é uma ciência sim, mas uma ciência que tem como uma de suas características ser poética, pois não pode ser reduzida a abstrações, a leis, a estruturas.

Seduzir ou simplesmente distrair, antes de tudo o ofício do campo é marcado pela necessidade de construir afirmações possíveis sobre o passado com os fragmentos encontrados. Para o historiador britânico E. H. Carr (1892-1982), o processo de reconstituição governa a seleção e interpretação dos fatos, são ações derivadas da mente do historiador imbuído das suas próprias experiências (CARR, 1978). Dessa forma, acontecimentos passíveis de serem qualificados para justificar a ação de um personagem ou grupo, supostamente impedem o historiador de ser totalmente neutro na escolha dos termos mais adequados para escrever a História. Como exemplo, podemos citar o movimento político-militar deflagrado no Brasil em 31 de março de 1964 que derrubou o governo constitucional do presidente João Goulart, e que tem sido exaltado como *revolução* ou condenado como *golpe* de Estado na sua denotação[57]. Qual dos dois termos é o mais apropriado? Por ser uma

[57] Enquanto os defensores e participantes do movimento de 31 de março referem-se a ele como Revolução de 1964, por considerar que o seu objetivo era produzir uma reformulação completa na vida política do país -

questão cujo consenso dificilmente poderá ser alcançado, isto leva a crer que o conhecimento histórico, neste sentido, carregará o tanto da parcialidade atribuída pelas suas categorias.

2.4.1 A noção de História Contemporânea do Brasil

O Brasil é uma nação que já nasce moderna, isto é, a história convencionalmente contada parte sempre do seu descobrimento nos primórdios do século XVI, não tendo, à semelhança da Europa e do Oriente, capítulos de uma História Antiga ou Medieval, embora seu território seja habitado desde tempos pré-históricos por povos indígenas. Os manuais e livros didáticos adotados no país costumam trazer a seguinte divisão para a História do Brasil pós-descobrimento: i) Colônia; ii) Império; iii) República.

O período colonial compreende o Descobrimento do Brasil no ano de 1500, passando pela designação de Reino Unido de Portugal em 1815, até a Declaração da sua Independência, em 1822. A partir daí a América Portuguesa se transformou em Império do Brasil, que teve seus três grandes momentos: o Primeiro Reinado, que vai de 1822 a 1831, o Período Regencial, entre 1831 a 1840, e o Segundo Reinado, de 1840 a 1889. Em 15 de novembro de 1889 se deu a Proclamação da República, que pôs fim à soberania do imperador D. Pedro II e deu início ao período republicano no Brasil.

Este último é demarcado por várias eras que compreendem marcos de transformação da história do país: a República Velha (1889-1930), a Era Vargas (1930-1945), a República Nova (1945-1964), o Regime Militar (1964-1985) e a Nova República (1985-atual).

O termo ‘contemporâneo’ nos remete sempre a algo que é do mesmo tempo, ou do tempo atual. Mas é importante ter claro, quando falamos de História, que essa fronteira é variável, não é a mesma para todos.

> “A existência, entre as fontes disponíveis, de testemunhos vivos pode ser um dos fatores considerados na delimitação cronológica da história do tempo presente, mas não pode ser o fator determinante. Talvez não se deva mesmo falar em fronteira, e sim reconhecer que há uma continuidade no tempo e que o presente revela fenômenos de longa duração. Ou que a

---

eliminando a corrupção e os mecanismos de poder que estariam sendo utilizados para favorecer a subversão comunista no Brasil - seus opositores e adversários definiram-no como golpe por tratar-se da deposição de João Goulart, presidente eleito legitimamente pelo povo. (DHBB, Portal do CPDOC/FGV, disponível em: http://cpdoc.fgv.br/acervo/dhbb/faq , acesso em 23/02/2011)

> história modifica a significação dos fenômenos à medida que as perspectivas mudam, e que o historiador busca no passado novos objetos de estudo em função das preocupações de seu tempo” (CPDOC, 2010).

Apesar de não haver uma delimitação precisa para o que chamamos de ‘Brasil Contemporâneo’ a ideia do termo empregado nesta pesquisa tem por influência direta a produção difundida pelo Centro de Pesquisa e Documentação de História Contemporânea do Brasil da Fundação Getulio Vargas (CPDOC/FGV), em especial o Dicionário Histórico-Biográfico Brasileiro. Esta é uma obra de referência que reúne conhecimento valioso sobre o país pós-1930, data em que Getúlio Vargas é colocado no epicentro do nosso cenário político e que, para muitos historiadores, marca o início de uma nova fase da história política brasileira.

O Brasil Contemporâneo é, então, o cenário que servirá como pano de fundo para o desenvolvimento da presente pesquisa. Mas, obviamente, as questões do domínio vão muito além deste pretenso recorte temporal e geográfico.

2.4.2 As velhas e novas questões do domínio

Pela impossibilidade de se alcançar de fato sua totalidade, a História sempre procurou se debruçar sobre histórias específicas inseridas dentro de unidades de sentido que conferem coerência a um corpo de documentos e a uma narrativa. Desde o século XIX, algumas unidades maiores têm predominado nos currículos escolares e literatura da área como os grandes contextos da História: uma civilização, uma sociedade, um estado, uma era (GUARINELLO, 2004). Ainda que a disciplina tenha passado por diversas mudanças metodológicas nos últimos dois séculos, os historiadores ainda circunscrevem suas Histórias dentro dessas unidades: fazem uma História da Roma Antiga, História da França Medieval, História do Brasil Contemporâneo ou História da Cultura Ocidental, por exemplo. São estas as grandes unidades de sentido, os objetos por excelência da História, dentro dos quais se recortam temas específicos: a economia, a política, a sociedade e assim por diante.

É verdade que algumas dessas grandes entidades assumem por vezes um caráter quase permanente, como categorias imutáveis, antropológicas. E ainda que o seu processo de fabricação esteja imbuído do cabedal pré-teórico e contemporâneo que conforma nosso modo

de ver e conceber o passado - num volver anacrônico -, não é possível pensar a História sem elas, sob pena de termos apenas uma sucessão de fatos soltos e desconexos.

A importância por fatos e datas, sem grandes análises de estrutura e conjuntura, prevaleceu na historiografia moderna até o início do século XX e orientou o ensino da disciplina no Brasil em todos os níveis até pelo menos a década de 1970 (CARDOSO, 1997). Era uma visão positivista que pretendia o resgate do passado 'tal como aconteceu', ou seja, se apoiando em fatos, grandes nomes e heróis, e privilegiando as fontes documentais primárias. Não à toa, esta tendência se fez acompanhar por um grande esforço de coleta e organização dos acervos de arquivos públicos, conjugando o interesse pelos feitos dos grandes homens e a história de caráter 'oficial', ou seja, vista de cima, pelas elites dirigentes (BURKE, 1992).

Contrário a esta abordagem, Marc Bloch (1886-1944) escreve que a história não é a simples acumulação dos acontecimentos, e nem estes o 'centro' do mundo. Antes, é preciso entender os fatos ao longo do tempo, estabelecendo suas relações de causa e efeito (SCHWARCS, 2001). O autor vê a História como uma 'ciência dos homens, no tempo': o tempo, na mesma medida em que é, por natureza, um 'continuum', traz também uma perpétua mudança.

Assim, a partir de meados do século XX percebe-se uma tendência, que logo predominaria, a despersonalizar a ação transformadora e a cadenciar mais lentamente os ritmos de mudança, como forma de superar as limitações dessa velha história política e positivista, para abrir-se às contribuições de outras áreas como a Sociologia, Antropologia e Economia, e para a influência do marxismo[58] (GUARINELLO, 2004). Povos, nações, estados e heróis cedem o lugar no palco para outros agentes, mais impessoais. Passam a predominar as grandes estruturas sociais e econômicas, que parecem mais reais e determinantes do que os acidentes de conjuntura, e a História começa a ser abordada pelos seus longos períodos em detrimento dos episódios, e fenômenos inscritos na longa duração tornam-se mais significativos do que os movimentos de fraca amplitude. Os instrumentos de análise passam a consistir de fontes seriais, estudos quantitativos e identificação de constantes sociológicas (FERREIRA, 2003). Aqui um parêntese sobre a sutil distinção que se faz entre História e histórias, ou seja, existe uma História, com H maiúsculo e existem apenas histórias, diversas,

[58] A concepção marxista vai preconizar o estudo da sucessão descontínua da história, especialmente no que se refere aos diferentes modos de produção para analisar como as riquezas são produzidas e distribuídas no interior de uma formação social, na qual a ação coletiva dos homens (luta de classes) constitui o motor da história (CARDOSO, 1997)

casuais e distintas entre si. Quando o historiador francês Fernand Braudel (1902-1985) escreve sobre o mundo mediterrâneo na época de Filipe II, ele claramente faz uma apologia da longa duração em detrimento da curta duração, dos eventos. Para o autor, ao que parece, os eventos seriam como os vaga-lumes a brilhar durante a noite, ao passo que esta, longa, densa e contínua, equivaleria à História (SILVA, 2009).

Mas um novo dilema da historiografia contemporânea vem ligar-se, de modo evidente, à crise dessas grandes estruturas. Da política à sociedade e desta à economia, o espaço da ação humana tornou-se cada vez mais restrito, menos eficaz. A capacidade de projetar um futuro e de conceber o passado como sua origem e sua causa, como a explicação prévia de um projeto a se concretizar, esvaeceu-se nas últimas décadas. Essa crise das grandes narrativas foi impulsionada pela 'nova história' francesa - originária da chamada *École des Annales* -, que implicava a defesa de uma concepção da história baseada em 'novos objetos', como os marginalizados, as mentalidades coletivas, as estruturas sócio-econômicas (BURKE, 1992; GUARINELLO, 2004). Em consequência, os objetos da História se multiplicaram, não apenas pela introdução de novos atores sociais (trabalhadores, imigrantes, mulheres, homossexuais), cujas histórias se tornaram relevantes, mas pelo recurso cada vez mais intenso ao estudo de caso, ao detalhe, à micro-história que se afasta dos grandes contextos.

Da ação individual e pontual de grandes homens às incursões acionais coletivas, da ação impessoal e cíclica das estruturas econômicas, ao recorte microscópico dos novos objetos de análise, não é possível negar hoje a coexistência de diferentes concepções e modos de fazer História. Segundo Foucault (1972) as velhas questões de análise tradicional – que ligação estabelecer entre acontecimentos díspares? Que continuidade os atravessa ou que noção de conjunto eles formam? Pode-se definir uma totalidade ou é preciso limitar-se a reconstituir encadeamentos? – são substituídas, ou melhor, acrescidas, por interrogações de outro tipo, como: que estratos são passíveis de serem isolados uns dos outros? Que tipos de séries instaurar? Que critérios de periodização adotar para cada uma delas? Que sistema de relações (hierarquia, dominância, escalonamento, determinação, causalidade) pode ser descrito entre uma e outra? Em que quadro de cronologia ampla pode ser determinado sequências distintas de acontecimentos?

Independente dos pressupostos – positivistas, marxistas ou de uma *Nouvelle Histoire* – existentes por trás destas e outras questões, o importante é que ao submeter a História a um tratamento filosófico, buscam-se os seus fundamentos últimos e o fio condutor que dará

inteligibilidade às miríades de acontecimentos aparentemente desconexos. Ou seja, identifica-se por trás das várias histórias particulares o elemento unificador, a regularidade; elabora-se uma explicação para a história e explicita-se o seu sentido. Os conceitos mais representativos do domínio se moldam, então, às respostas para as velhas e novas questões filosóficas.

2.4.3 Os métodos e as abordagens na história

Os métodos na História sofrem então de todas as influências decorrentes de grupos, nem sempre convergentes, que seguem suas próprias convicções no que tange aos objetos de estudo, suas escolhas e formas de abordagem.

Atento ao que diz respeito ao relato e à investigação histórica, Bloch (2001) conclama que "todo livro de história digno desse nome deveria incluir um capítulo ou uma sequência de parágrafos que se intitularia*: Como posso saber o que vou dizer*?" Resgataria-se, assim, a própria história da investigação, com seus problemas e métodos.

O autor aponta que é indispensável ao historiador possuir ao menos um verniz de todas as principais técnicas de seu ofício, técnicas estas que irão ajudá-lo no tratamento das questões a serem levantadas diante dos documentos e do discurso histórico. Da mesma forma, Le Goff afirma que "a História só é feita recorrendo-se a uma multiplicidade de documentos e, por conseguinte, de técnicas", reassegurando o que dizia Bloch: "poucas ciências são obrigadas a usar, simultaneamente, tantas ferramentas dessemelhantes" (LE GOFF, 2001). Segundo ele, isto se dá porque os fatos humanos são, em relação a todos os outros, complexos. E porque o homem se situa na ponta extrema da natureza.

Quanto aos modos de fazer História, adotam-se diversas contribuições de técnicas advindas de algumas das especializações do campo, como a História Oral, a História Serial e Quantitativa, a Micro-História e a História Regional. Segundo Barros (2006), a História Oral recorre a um tipo de fonte produzido pelo próprio historiador ao trabalhar com os testemunhos orais: "... suas preocupações neste âmbito estarão relacionadas ao tipo de entrevista que será utilizado na coleta de depoimentos, ao uso ou não de questionários pré-direcionados, e assim por diante" (BARROS, 2006, p. 133). Já a História Serial, introduzida em meados do século XX, está intimamente associada à história quantitativa, à séries históricas que apresentam determinado tipo de homogeneidade capazes de serem analisadas sistematicamente pelo historiador. Quanto à Micro-História, esta leva à redução da escala de observação do historiador, a fim de capturar em uma sociedade aquilo que habitualmente escapa àqueles que

preferem trabalhar com um ponto de vista mais panorâmico, mais generalista ou mais distanciado. Por outro lado, examinando um espaço de atuação onde os homens desenvolvem suas relações sociais, políticas e culturais, a História Regional resgata um tipo de saber historiográfico que permite estudar uma ou mais dimensões neste espaço, tanto no que diz respeito ao seu desenvolvimento interno, quanto ao que se refere à inserção em universos mais amplos.

A História se repartiu em objetos e temáticas que não são uniformes nem equivalentes, e que mudam de abrangência ou de sentido de acordo com os interesses de quem financia, produz ou consome seus relatos.

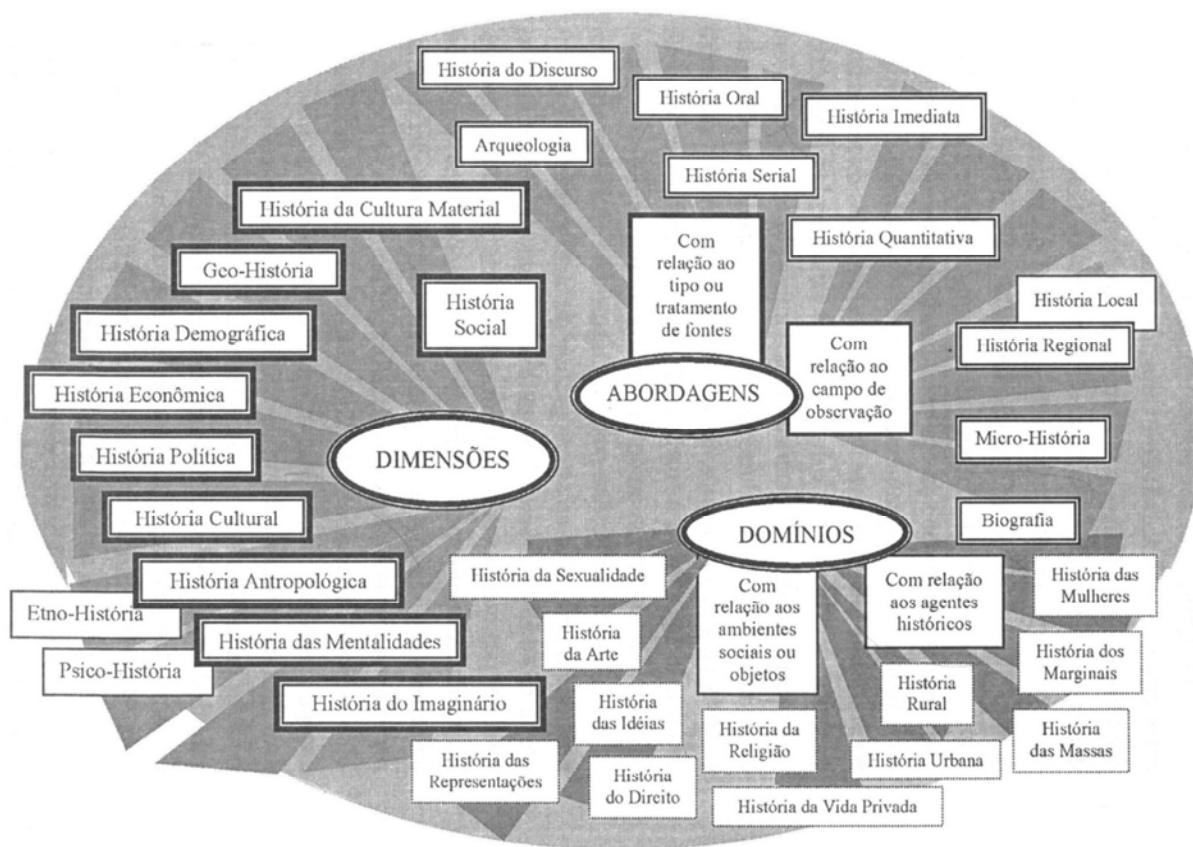



Figura 28 – Domínios, dimensões e abordagens na História
(Fonte: BARROS, 2006)

A figura acima é uma amostra de domínios que podem ser investigados, tendo em vista as diferentes dimensões históricas e métodos de abordagem existentes. Faz-nos pensar a História como um campo de posibilidades, onde as interconexões podem estar associadas a uma reflexão que endereça à importância dos conceitos, a relatividade das "divisões" e a necessidade de buscar um embasamento teórico para cada escolha adotada.

2.4.4 Os documentos na história

Segundo Bloch (2001), mesmo o mais claro e complacente dos documentos não fala senão quando se sabe interrogá-lo. Ou seja, é a pergunta que fazemos que condiciona a análise e, no limite, eleva ou diminui a importância de um texto retirado de um momento afastado. Segundo ele, nenhum objeto tem movimento na sociedade humana exceto pela significação que os homens lhe atribuem, e são as questões que condicionam os objetos e não o oposto. Esse tipo de visão crítica – oposta aos modelos mais tradicionais de historiografia, que acreditavam naquilo que Le Goff chamou de "imperialismo dos documentos" – marcou Bloch e toda a primeira geração dos Annales. Tratava-se de uma espécie de guerra de 'trincheiras' contra a história exclusivamente política e militar, uma história até então segura e tranqüila diante dos eventos e da realidade que buscava anunciar (SCHWARCS, 2001).

Sob uma abordagem antropológica que toma os documentos como fonte histórica, Heymann (2010) destaca que a representação que se faz destes está intrinsecamente associada a categorias como 'memória', 'patrimônio' e 'legado', que lhes conferem sentido e apontam para seus usos e estratégias de apropriação. Já Murguia (2010) remete à questão do documento como dispositivo de produção da verdade: "eles atuam e têm serventia porque são verdadeiros no sentido da sua indicialidade enquanto vestígios". E Freitas (2010) aponta a existencia de correntes historiográficas que consideram documentos como sendo tudo aquilo que seja incorporado como fontes para construção de narrativas históricas.

Documentos produzidos a partir de eventos históricos podem ser basicamente de dois tipos: primários, quando se constituem fontes diretamente ligadas ao evento em si (por exemplo, o inquérito policial, a carta-testamento de Getulio Vargas); ou secundários, quando produzidos para narrar e explicar o que aconteceu (textos, artigos, verbetes). Dessa forma, documentos são coligidos como prova de autenticidade e evidência, além de constituirem elementos adicionais de informação.

Uma ontologia que versa sobre o domínio da história serve-se dos documentos sob estes dois aspectos.

2.4.5 A análise do discurso histórico

Um fator importante da análise histórica é o vocabulário, a terminologia, ou ainda, a 'nomenclatura'. Marc Bloch discute como o historiador deve conduzir sua análise com o auxílio de uma dupla linguagem, não só a da época estudada, o que lhe permite evitar o

anacronismo, mas também a do aparato verbal e conceitual da disciplina histórica atual (LE GOFF, 2001): “Estimar que a nomenclatura dos documentos possa bastar completamente para fixar a nossa seria o mesmo, em suma, que admitir que eles nos trazem a análise toda pronta (BLOCH, 2001).” Le Goff (2001) conclama que, ainda que o historiador não tenha o fetichismo da etimologia (‘uma palavra vale bem menos por sua etimologia do que pelo uso que se faz dela’), ele deve se dedicar ao estudo dos sentidos, à ‘semântica histórica’.

Quando falamos em efetuar a análise semântica de um material histórico que instanciará um domínio, significa também descrever o processo discursivo do qual este material é a realização. É importante um quadro teórico que hierarquize os fenômenos, com um corpo de conceitos articulados. Centrada na palavra, a semântica histórica, aqui mais exatamente a lexicologia histórica, se constrói sob um vasto programa de investigação: especificidade do vocabulário, relações deste vocabulário com o meio social contemporâneo e o lugar deste vocabulário no corpus verbal da língua vernácula (ROBIN, 1977).

Ao abordar o campo da história do discurso, Barros (2006) aponta que os textos podem ser pensados como ‘objeto de significação’ e ‘objeto de comunicação’. No primeiro caso, considera-se a organização interna do texto e no segundo, a reflexão sobre o contexto histórico-social em que está inserido:

> “De acordo com esta visão complexa e multidimensional do texto que se mostra a mais adequada para o historiador, podemos dizer que a análise de um discurso deve contemplar simultaneamente três dimensões fundamentais: o *intratexto*, o *intertexto* e o *contexto*. O ‘intratexto’ corresponde aos aspectos internos de texto e implica exclusivamente na avaliação do texto como objeto de significação, o ‘intertexto’ refere-se ao relacionamento de um texto com outros textos; e o ‘contexto’ corresponde à relação do texto com a realidade que o produziu e o envolve”. (BARROS, 2006, p. 136-137).

A Linguística contribui ao permitir substituir o dado do texto por uma lógica do texto. Serve para revelar a economia interna de uma ideologia, mas nunca para estabelecer sua função social. Ou seja, ela descreve o texto, identifica precisamente o que diz o texto, seu arranjo interno, mas não nos dá sua ideologia, sua significação social.

Na análise do discurso histórico percebemos também que um dos seus grandes problemas se refere à enunciação: onde começa, onde acaba a enunciação? Como lidar com a intervenção do sujeito da enunciação na constituição da significação do discurso? Percebemos que a afirmação aparentemente mais neutra pode remeter a uma intenção, e talvez designar de maneira intuitiva e paralela, o lugar que ocuparia a noção de 'sistema de representações' ligada à noção de ideologia, no quadro do materialismo histórico[59].

Para Régine Robin (1977) a questão principal seria: uma análise do discurso pode permanecer intralinguística, sem considerar as condições de produção desse discurso, o quadro institucional no qual ele é produzido, as relações de força que presidem o momento do discurso, as formações ideológicas que governam este ou aquele tipo de retórica?

2.4.6 Os desafios da modelagem conceitual no domínio da História

Esta pesquisa se propõe a investigar as primitivas de uma ontologia que possa estender-se para uma base de conhecimento. Nesse sentido, os desafios aproximam-se mais das questões de identificação e posicionamento adequado dos conceitos em um sistema de representação capaz de lidar com o discurso histórico, seja qual for a interpretação que lhe é atribuído, e mesmo, que consiga lidar com essas interpretações.

Podemos sintetizar os desafios nas seguintes questões:

- Quais seriam as categorias principais no domínio da História?
- Quais os limites temporais para ações, eventos e processos? O que distingue um do outro? O conceito de evento para a historiografia seria o mesmo que para as outras ciências?
- Quais os critérios para a existência de identidade das entidades do domínio? O que garante esta identidade? (GUARINO, 1997; GUIZZARDI, 2005)?
- Como ordenar os eventos?
- Tempo e espaço são grandezas absolutas, universais e independentes, ou grandezas existencialmente dependentes de outras entidades?
- Como lidar com a incerteza na datação de eventos?
- Como lidar com interpretações na história?

---

[59] O principal objetivo do materialismo histórico é explicitar as causas que governam a direção da mudança histórica. Para Marx, a base da mudança é material: é a economia, o trabalho do homem.

Smith (2002), ao discutir a relação entre ciência (no caso, a História) e Ontologia (enquanto domínio de conhecimento preocupado com a existência dos seres), afirma que enquanto o papel da ciência é ‘explicar’ a natureza, o papel da Ontologia seria, uma vez que fossem obtidas as explicações, descrever, organizar e sistematizar o conhecimento obtido. Esta é a orientação que fundamenta o presente trabalho. Trabalho este que reconhece, acima de tudo, que para muitas das questões acima não há simplesmente uma única resposta possível, mas sim formas diferentes de apropriação do ferramental técnico e metodológico disponível, capaz de endereçá-las.

3 Metodologia de pesquisa

A metodologia da pesquisa se baseia em uma abordagem prática, uma vez que será proposta uma ontologia inicial amparada por pressupostos visitados. É também de caráter qualitativo, pois visa a coletar conceitos e suas inter-relações, a partir de diversas fontes de conhecimento sobre o domínio. Assim, segundo classificação proposta por Lakatos e Marconi (1991), trata-se de uma pesquisa aplicada quanto à sua natureza, qualitativa quanto à forma de abordar o problema e exploratória quanto aos seus objetivos.

Sobre os procedimentos de pesquisa, algumas direções foram demarcadas. Primeiramente, a pesquisa bibliográfica que permitiu delimitar o espaço de debate envolvendo os conceitos fundamentais acerca do tema. Buscaram-se principalmente autores seminais e formuladores das teorias, embora estejam presentes, conforme o caso, autores comentadores dos primeiros.

Foi realizada revisão de literatura sobre metodologias de construção de ontologias cujos critérios de seleção se basearam na clareza quanto à sistematização das etapas, existência de documentação e citações na área. Foram analisadas as metodologias de Gruninger e Fox (1995), a Methontology de Gómez-Pérez et al (1996, 1997), o método 101 de Noy e McGuinness (2001) e a Ordnance Survey (KOVACS et al, 2006), envolvendo as seguintes etapas: especificação, aquisição do conhecimento, conceitualização, integração com outras ontologias, formalização, implementação e documentação. Estas três últimas não chegaram de fato a serem aplicadas neste estudo, porém suas diretrizes são explicitadas para fins de entendimento do processo como um todo.

Fechando o marco teórico, foi feita uma análise do domínio da História, onde são expostas as principais questões que permeiam o ofício do historiador nas palavras de alguns autores expoentes do campo historiográfico, como Marc Bloch, Jacques Le Goff, Peter Burke, Ciro Flamarion Cardoso e Marieta de Moraes Ferreira.

O levantamento de ontologias e projetos afins faz parte da metodologia da pesquisa, pois questões sobre reuso e interoperabilidade são centrais no processo de construção do modelo. Este levantamento está descrito no próximo capítulo, e seu desdobramento se traduz na modelagem conceitual da ontologia em si, sistematizada como o campo empírico desta dissertação. A experiência consiste em aplicar o que foi aprendido com as metodologias estudas, descrevendo as seguintes etapas consideradas mais apropriadas:

- Identificação do domínio, escopo e questões que a ontologia deverá responder;
- Identificação de ontologias que possam servir de referência;
- Identificação dos conceitos recorrentes ao domínio com a ajuda de textos, verbetes e listas de vocabulário controlado;
- Identificação das categorias principais, a partir do agrupamento coerente dos conceitos;
- Identificação dos atributos e relacionamentos das entidades do domínio.

Os resultados buscam responder às questões centrais da pesquisa - expostos no capítulo introdutório -, que versam sobre quais princípios metodológicos orientariam o desenvolvimento de uma ontologia no domínio da História; como as estruturas ontológicas representariam apropriadamente a dinamicidade temporal presente na História; e quais conceitos deveriam ser integrados à ontologia.

A partir desse raciocínio, e com base na literatura, foram propostas algumas classes nucleares para o domínio em questão, com uma seleção de possíveis atributos e relacionamentos concebidos principalmente a partir do cotejamento com ontologias externas. Para cada classe, foram expostas definições segundo sua inserção no campo histórico, e ao final, um quadro síntese com uma seleção das propriedades levantadas.

4 Estado da arte e ontologias relacionadas

A história como disciplina e ramo do conhecimento trata essencialmente de registrar e explicar eventos relevantes do passado (Merriam-Webster Dictionary; CARR, 1978; SEWELL JR, 1996) e a importância dos eventos é recorrente nos relatos de experiências de desenvolvimento de ontologias focadas em, ou próximas a este domínio (NAGYPÁL, 2005; HYVOBEN et al, 2007).

Para obter um panorama do estado da arte dessas experiências, o presente capítulo reúne algumas ontologias que lidam com a noção de evento sob algum aspecto. Ontologias de topo ou de domínio, elas foram selecionadas principalmente pela relevância apontada na literatura, nos estudos de caso e existência de documentação. Serão examinadas no que as fazem aproximar-se, distanciar-se ou contribuir de alguma forma para a ontologia de história ora proposta, doravante identificada como Ontologia de História ou Ontologia HIST.

Das ontologias visitadas, são consideradas de topo ou de fundamentação OpenCYC, DOLCE e Unified Foundational Ontology. As que tratam eminentemente de eventos são Event Ontology, Dolce Ultralite Ontology, Simple Event Model e Linking Open Description of Events. Já a CIDOC-CRM e Europena Data Model não chegam a ser identificadas como ontologias de topo, no sentido de abrigarem noções gerais do conhecimento, mas não são exatamente ontologias de domínio visto que seus objetivos estão em servirem de *frames* à representação de conceitos na área da cultura e patrimônio, de forma abrangente; elas se situam no meio termo. VIVO e VICODI são experiências mais restritas a domínio: a primeira no campo acadêmico e de publicações, e a segunda na história da Europa moderna. A DAML Time Ontology é uma ontologia de topo focada na modelagem de noções temporais independentes de domínio.

4.1 OpenCYC

OpenCYC é um subconjunto da Cyc, ontologia de topo bastante referenciada e constituída de centenas de milhares de conceitos e asserções extraídos da realidade. A Cyc se autodenomina ‘uma base de conhecimento do senso comum’ destinada a prover entendimento legível para programas computacionais”[60]. É implementada na linguagem CycL – cuja sintaxe deriva da lógica de primeira ordem (lógica formal) e do LISP (lógica matemática) –, e

[60] http://www.cyc.com/technology/whatiscyc

apoiada por um subsistema de processamento de linguagem natural da qual fazem parte o léxico, um *parser* sintático e um interpretador semântico (para navegar e realizar inferências). Em suma, é uma base de conhecimento complexa que agrega várias ontologias de domínio (chamadas microteorias), e da qual OpenCyc é uma versão enxuta e *opensource*.

Conforme a figura 23 mostra[61], `Event` e `StaticSituation` são classes disjuntas subordinadas à classe `Situation`, e visam denotar as duas situações possíveis para uma dada realidade: mudança e persistência.

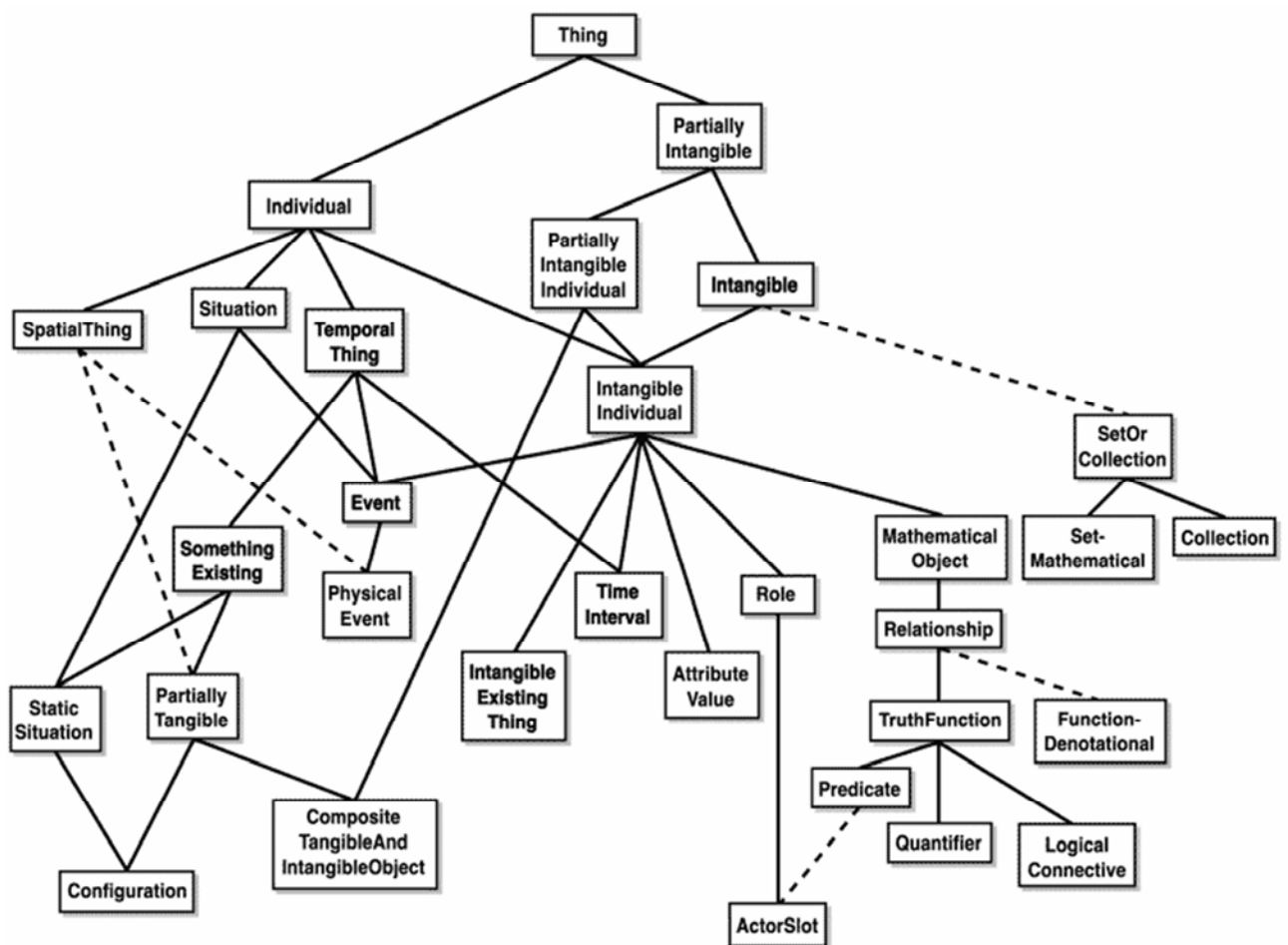



Figura 29 - Visão parcial da ontologia OpenCYC. As linhas tracejadas representam relações indiretas.

`Event` e `StaticSituation` também são especializações de `TemporalThing`, que é a classe que reúne todas as coisas passíveis de possuírem uma localização ou extensão temporal. Em Cyc, não há distinção entre eventos e ações, ou seja, tudo que possa ser dito

[61] Figura extraída de http://www.eaft-aet.net/fileadmin/files/VAKKI/christophe_roche.ppt. Acesso em 20/06/2012.

como ‘acontecendo’ é um evento, desde o vôo de um pássaro até uma pessoa estar pensando ou aprendendo.

Pelo que é possível depreender da documentação oficial[62], eventos se ligam a outros eventos conformados pela propriedade `role`, e podem ser decompostos nas suas partes temporais ou espaciais a partir da propriedade `subEvent`. Cada subevento é parte de um dado evento, e deve estar completamente contido nele, pelo aspecto temporal ou espacial (relação todo-parte).

4.2 DOLCE Ultralite Ontology (DUL)

A DUL Ontology[63] é uma combinação da ontologia DOLCE (*Descriptive Ontology for Linguistic and Cognitive Engineering*) e DOLCE DnS (*Descriptions and Situations*) em uma versão mais simples e intuitiva. A primeira, DOLCE, foi desenvolvida no *Laboratory for Applied Ontology* em Trento, por Nicola Guarino, e a segunda, uma especialização da primeira acrescida de aspectos da realidade social, por Aldo Gangemi no *Semantic Technology Lab*, em Roma, também na Itália. As questões de competência elencadas pela ontologia DUL envolvem em alguma medida aspectos relativos a eventos (*what participates in what event? what are the entities in the setting of that situation? what description is satisfied by what situation?*[64]). Mas antes de apresentá-la, é importante introduzir aquela que a originou.

DOLCE é descrita pelos autores como sendo uma ontologia de particulares. A distinção ontológica fundamental entre universais e particulares neste contexto é quanto ao aspecto de instanciação: particulares são entidades que não possuem instâncias e universais possuem (MASOLO et al, 2003). Universais aparecem na ontologia dos particulares para caracterizá-los, correspondendo às propriedades e relacionamentos existentes entre eles. Além disso, DOLCE é fortemente orientada pela distinção entre endurantes e perdurantes no que se refere à localização das entidades no tempo; enquanto os primeiros estão completamente presentes em qualquer tempo que existam (uma pessoa, um carro), os últimos estão apenas parcialmente presentes porquanto são constituídos de ‘partes temporais’ (um processo, um evento) (GANGEMI et al, 2002).

---

[62] http://www.cyc.com/cycdoc/vocab/top-vocab.html
[63] http://www.loa-cnr.it/ontologies/DUL.owl#Event.
Mais informação em: http://en.wikipedia.org/wiki/Upper_ontology_(information_science).
[64] http://ontologydesignpatterns.org/wiki/Ontology:DOLCE+DnS_Ultralite

A taxonomia das categorias mais básicas dos particulares assumidos em DOLCE pode ser visualizada na figura a seguir.

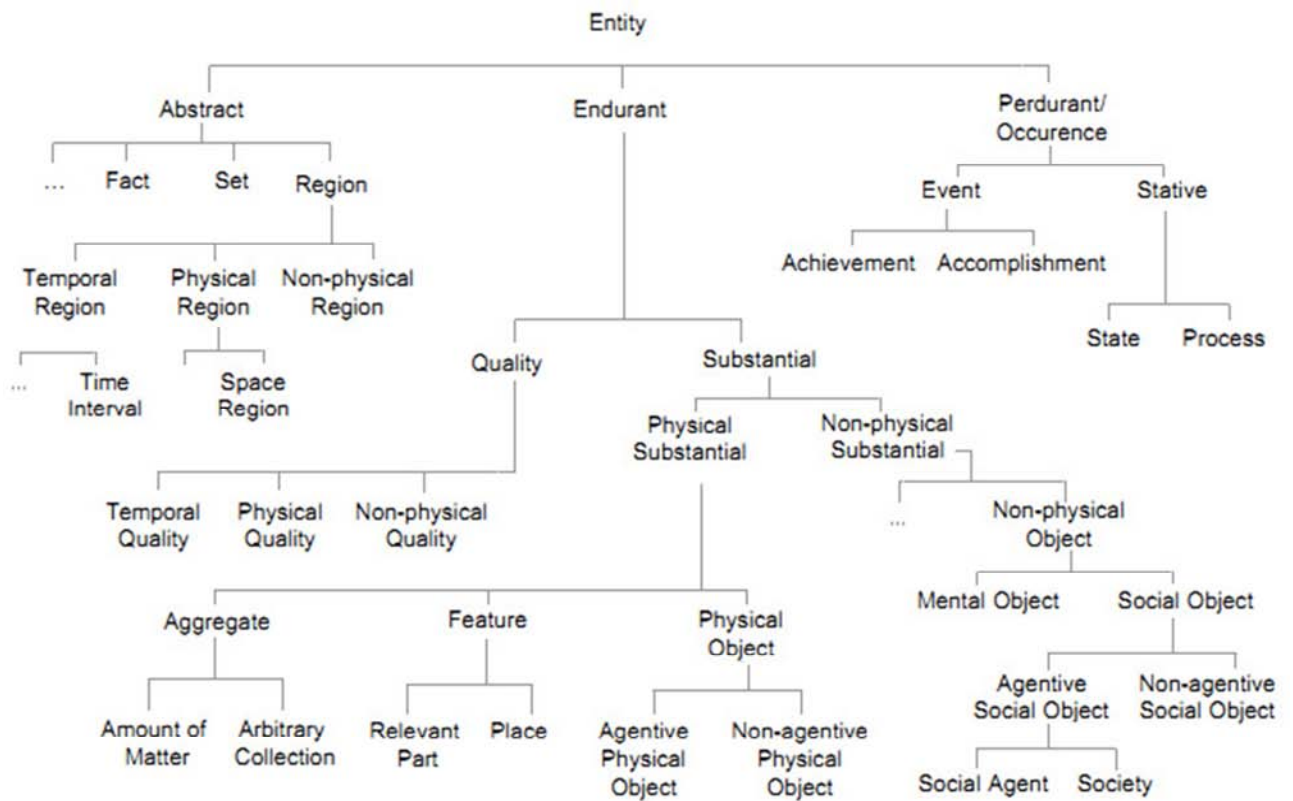



Figura 30 - Taxonomia das categorias básicas de DOLCE. (Adaptado de GANGEMI et al, 2002)

Baseada em DOLCE, a ontologia DUL tornou os nomes das classes mais intuitivos, simplificou as relações temporais e espaciais, enxugou algumas subordinações e manteve as noções de qualidade e abstração como primitivas. As classes principais da ontologia englobam `Agent`, `Object`, `Event`, `InformationEntity`, `Abstract` e `Quality`, e são todas subordinadas a `Entity`. A figura a seguir é uma ilustração criada a partir da documentação existente sobre a DUL.

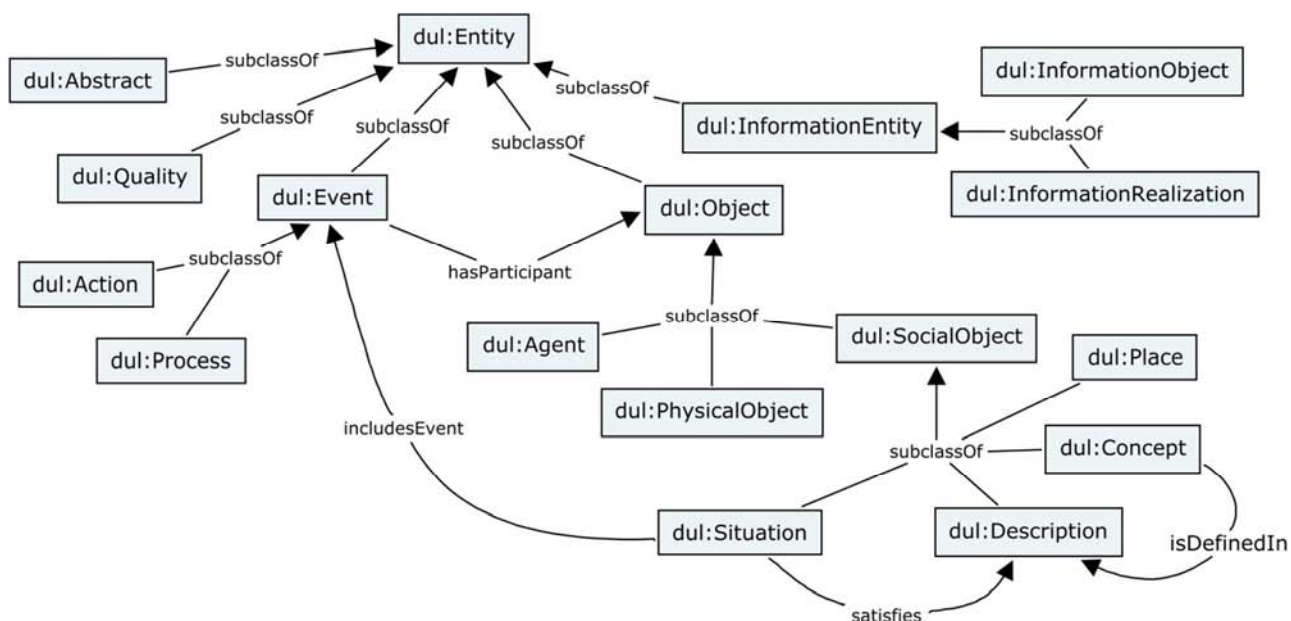



Figura 31 - Visão parcial da Ontologia DUL
(criado a partir do OWL disponível em: http://www.ontologydesignpatterns.org/ont/dul/DUL.owl)

`Event` e `Object` são classes disjuntas, usadas para representar entidades que existem no tempo e entidades que existem no espaço, respectivamente. Estas últimas, aliás, abrangem coisas imaterias e abstratas, como os conceitos mentais. `Qualidade` é a característica atribuída a um objeto ou evento, e seu valor pode ser localizado no escopo da classe `Abstract`.

O desejo de separar eventos das suas interpretações é o que parece mover a abordagem da DUL. Teoricamente falando, eventos de um modo geral podem ser classificados tendo por base seu 'aspecto' (estático, contínuo, realizado), sua 'causalidade' (intencional, natural) ou seus 'participantes' (humano, objeto físico, abstrato), resultando em diferentes visões ao mesmo tempo de um evento em particular (W3C, 2011). A ontologia DOLCE Full utiliza esta abordagem para classificar eventos, o que parece complicado para o domínio da história porque ela apenas cria identidades diversas para o mesmo evento segundo os aspectos escolhidos de observação. Esta é uma questão muito presente na vertente filosófica da área de representação do conhecimento (SOWA, 2000; SMITH, 2004; SMITH e CEUSTER, 2010; MERRIL, 2010), que vai determinar o tipo de compromisso ontológico desejado para o modelo conceitual a ser concebido. Temos de um lado os realistas defendendo que uma ontologia é determinada de acordo com a sua conformidade com o mundo real, que independe da cognição humana e pode ser aferida de forma objetiva (ou seja, eventos não mudam pela maneira como os descrevemos); e de outro, os construtivistas com a percepção de que,

qualquer que seja a ‘realidade verdadeira’, ela não pode ser modelada sem levarmos em consideração os aspectos observáveis e descritivos possíveis.

Conforme a documentação relata, a classe `dul:Event` se refere a qualquer processo (físico, social ou mental), estado e, claro, evento. A diferença é que em DUL, várias instâncias de eventos podem participar de uma dada situação, representada na ontologia pela classe `dul:Situation`. Esta, por sua vez, satisfaz certa descrição localizada em `dul:Description`. Dessa forma, uma situação (que pode ser até mesmo um evento) tem a possibilidade de ser descrita (ter várias interpretações e pontos-de-vista) a partir dos diversos eventos ligados a ela e permanecer singular, ou seja, com a mesma identidade.

Por ser um modelo flexível o bastante para acolher as diferentes realidades daquilo que se deseja representar, os princípios adotados pelo DUL mostram-se bastante interessantes.

4.3 CIDOC CRM - CIDOC Conceptual Reference Model

No âmbito do patrimônio cultural, o CIDOC Conceptual Reference Model[65] (CRM) é uma ontologia de referência que fornece uma estrutura conceitual em um alto nível de generalidade. É um padrão ISO (ISO 21127:2006) e seu modelo visa apoiar a organização e representação da informação relativa a bens culturais, principalmente no escopo de curadoria de museus. São cerca de 90 classes e 150 propriedades concebidas para representarem não apenas as características intrínsecas dos objetos, mas também seus ciclos de vida. A um primeiro nível são encontradas as classes: `Persistent-Item`, `Time-Span`, `Place` e `Temporal-Entity`.

`Event` é uma entidade temporal e, segundo a documentação oficial (ICOM/CIDOC, 2011, p. 4), compreende “mudança de estado em sistemas culturais, sociais ou físicos, independentemente da escala, provocada por uma série ou grupo de fenômenos coerentes físico, cultural, tecnológico ou jurídico”, afetando diretamente instâncias da classe `Persistent-Item` (`Actor` e `Thing`). A classe `Event` fornece as subclasses: `Production`, `Creation`, `Dissolution`, `Acquisition` e `Curation`, e suas instâncias são modeladas de acordo com delimitações espaços-temporais providas por `Beginning-of-Existence` e `End-of-Existence`. As possíveis propriedades para eventos são `had-participant` e `ocurred-in-the-presence-of`.

---

[65] http://cidoc.mediahost.org/standard_crm%28en%29%28E1%29.xml

O esquema a seguir apresenta uma visão parcial do CIDOC-CRM, com foco na entidade `Event`:

Figura 32- Visão parcial da classe Event e seus relacionamentos (ICOM/CIDOC, 2011)

Além de `Event`, a hierarquia CIDOC coloca sob a classe `Temporal-Entity`, a classe `Condition-State` para descrever o estado ou a prevalência de certa condição física de um objeto.

CIDOC CRM Class Hierarchy

E1	CRM Entity
E2	- Temporal Entity
E3	- - Condition State
E4	- - Period
E5	- - - Event
E7	- - - - Activity
E8	- - - - - Acquisition Event
E9	- - - - - Move
E10	- - - - - Transfer of Custody
E11	- - - - - Modification

Figura 33 - Visão parcial da hierarquia de classes CIDOC CRM (ICOM/CIDOC, 2011)

Esta separação entre estado e mudança das coisas aproxima-se do esquema proposto pela ontologia OpenCYC que, para representá-los, fornece as classes `openCyc:StaticSituation` e `openCyc:Event`, subordinadas a `openCyc:Situation`..

4.4 Europeana Data Model (EDM)

Inspirado na biblioteca de Alexandria que desejava acolher todo o conhecimento do mundo, o projeto Europeana, levado a cabo pela União Européia, tem a ambição de congregar e tornar acessível o imenso patrimônio cultural de domínio público das bibliotecas nacionais do continente europeu, como livros raros, antigos ou cujas edições se esgotaram, pinturas, músicas, manuscritos, mapas. Para tanto, é necessário que os metadados que descrevem os milhares de itens digitais espalhados pelas instituições sejam reunidos e possam interoperar entre si através de uma camada de semântica compartilhada.

Esta camada é provida pelo Europeana Data Model (EDM)[66], um modelo de representação que adota o reúso de vocabulários e ontologias existentes, como o Dublin Core, SKOS e OAI-ORE, adaptando-os ao contexto da Europeana.

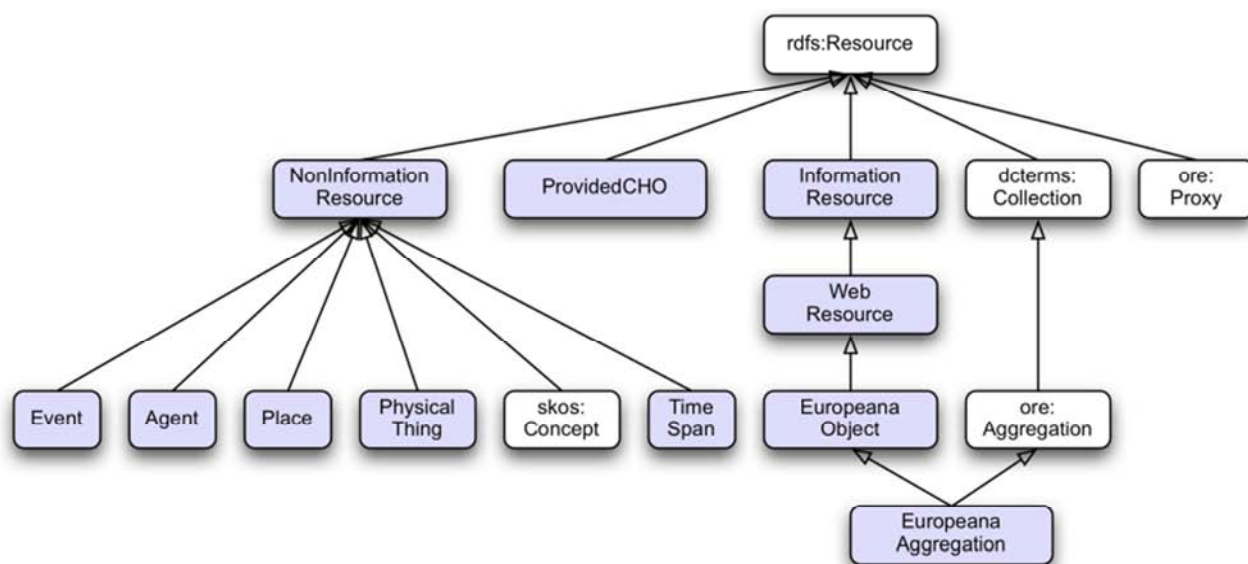



Figura 34 - Europeana Data Model. As classes em azul são introduzidas pelo EDM, e as de cor branca vêm de outros esquemas (Fonte: EDM Documentation)

A classe `edm:Event` é diretamente subordinada à classe que representa todos os recursos não informacionais da ontologia, e está no mesmo renque de `edm:Agent`, `edm:Place`, `edm:PhysicalThing`, `edm:TimeSpan` e `skos:Concept`, este último um

[66] http://pro.europeana.eu/edm-documentation

conceito emprestado do SKOS e usado para contextualizar os recursos da Europeana, podendo ser uma ideia, um pensamento ou noção abstrata.

A documentação fala que eventos no EDM têm a mesma conotação apresentada na ontologia CIDOC-CRM, ou seja, abrangem as ocorrências possíveis para o ciclo de vida dos objetos informacionais - como produção e custódia. Mas ela vai um pouco além, estendendo-se para a representação de eventos históricos em geral. O objetivo é gerar um modelo capaz de localizar os objetos em contextos, ligando-os aos eventos do passado, pessoas, lugares, etc, constituindo uma rica base de conhecimento de caráter eminentemente histórico.

4.5 Event Ontology (EO)

A ontologia Event[67], originalmente desenvolvida para a área da música, adota a concepção de Allen & Ferguson (1994) que diz que "eventos são a forma pela qual agentes classificam certos padrões relevantes de mudança". O modelo confere à entidade `Event` a noção de reificação (RAIMOND & ABDALLAH, 2007), imputando-lhe localização, tempo, agentes, fatores e produtos, conforme observado no modelo extraído da página oficial da ontologia:

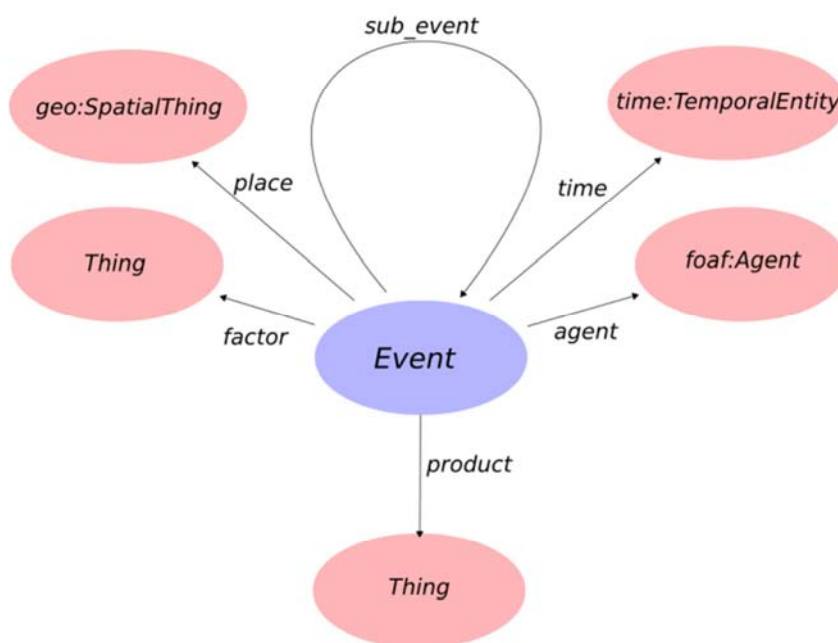



Figura 35 - Event Ontology (Fonte: EO Documentation)

[67] http://purl.org/NET/c4dm/event.owl#Event

As seguintes propriedades relacionam as entidades da ontologia: i) `hasAgent`, liga o evento a um agente ; ii) `hasFactor`, liga o evento a um fator, que pode ser um fenômeno ou uma causa abstrata; iii) `hasPlace`, liga o evento a uma localização; iv) `hasProduct`, relaciona o evento a algo produzido pelo evento, pode ser uma música, um objeto físico, etc.; v) `sub_event`, liga a outros eventos ou sub-eventos; vi) `time`, liga o evento a uma instância temporal.

Por ter uma estrutura simples, porém abrangente, Event tem suas classes emprestadas para diversas ontologias, como SNaP[68], BBC PO[69], BIBO[70], VIVO[71], etc.

4.6 VIVO Ontology (VIVO)

VIVO[72] é um projeto de web semântica voltado para a integração de recursos acadêmicos, que começou a ser desenvolvido em 2003 na Cornell University e hoje tem a participação de várias instituições de ensino superior nos EUA.

Figura 36 - VIVO Ontology (fonte: VIVO Documentation)

[68] Simple News and Press Ontologies. http://data.press.net/ontology/
[69] BBC Programmes Ontology. http://purl.org/ontology/po/
[70] Bibliographic Ontology. http://bibliontology.com/specification
[71] VIVO Ontology. http://purl.bioontology.org/ontology/vivo
[72] http://purl.bioontology.org/ontology/vivo

A ontologia criada visa representar pesquisadores no contexto das suas experiências, produção, interesses e filiações, e conterá basicamente informações sobre pessoas, organizações, publicações e cursos. Sua estrutura comporta entidades provenientes da BIBO, Dublin Core, Event Ontology, FOAF e SKOS. Uma visão parcial do modelo, em seu conjunto, é mostrada na figura 30.

A classe `Event` é tomada da ontologia Event, o que a aproxima da definição de Allen e Ferguson, como uma classificação arbitrária da região espaço-temporal determinada por um agente cognitivo. “Evento é algo que acontece em um dado lugar e tempo, possuindo os seguintes relacionamentos adicionais: agentes ativamente participantes, fatores passivos e produtos de trabalho” (BioPortal, 2011). VIVO fornece as propriedades `dateTimeInterval` e `description` para contextualizar os eventos, e restringe seu escopo (*range*) aos seguintes tipos de evento: *Competition, Conference, Course, Exhibit, Hearing, Interview, Meeting, Performance, Presentation, Workshop.* Eventos podem ocorrer também em espaços virtuais e serem divididos em séries (sub eventos): série de conferências, série de apresentações, etc.

4.7 Simple Event Model (SEM)

O Simple Event Model[73] foi desenvolvido com o intuito de representar e compartilhar informação baseada em eventos sem ter um domínio específico em mente, mas a inspiração vem da modelagem de eventos derivados de fatos na história (VAN HAGE et al, 2011). De acordo com seus autores, o modelo possui uma estrutura de representação enxuta, porém suficiente para descrever, em um evento, quem fez o que, quando e onde, além dos papéis atribuídos a cada ator e o tempo durante o qual cada papel é válido.

A estrutura conceitual de SEM constitui-se dos seguintes elementos: um núcleo de classes principais, tipos para identificar indivíduos de cada classe e restrições. São quatro as classes, `sem:Event`, `sem:Actor`, `sem:Place` e `sem:Time`, e a cada uma delas está associada uma classe `sem:Type` para atribuição de tipos aos seus indivíduos. Indivíduos e tipos são em geral emprestados de outros vocabulários. Por exemplo, `sem:PlaceType` “capital” (tgn: 1000841) pode ser extraído do Getty Tesaurus of Geographical Names (TGN).

[73] http://labs.mondeca.com/dataset/lov/details/vocabulary_sem.html

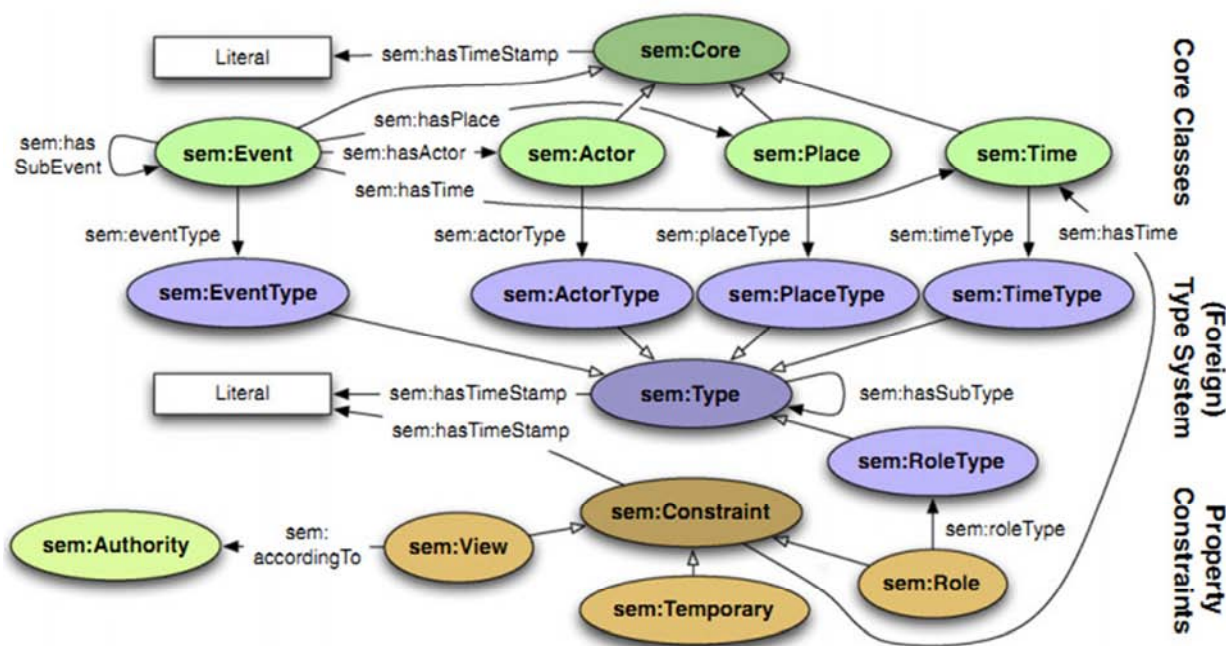



Figura 37 - Simple Event Model (SEM)
Extraído de VAN HAGE et al (2011)

Segundo os autores, as classes `sem:Role`, `sem:Temporary` e `sem:View` estabelecem as restrições (`sem:Constraints`) do modelo. Papéis (`sem:Role`) podem ser especificados para indivíduos de todas as classes principais, por exemplo, atores ('assassino', 'revolucionário') ou lugares ('capital'), e não são tipos dependentes (como em 'para ser mãe é necessário existir pelo menos um filho'). A classe `sem:Temporary` define os limites temporais nas quais certas propriedades se aplicam, incluindo a atribuição de papéis. Quanto ao `sem:View`, trata-se de uma classe interessante usada para definir pontos de vista e opiniões. Por exemplo, o movimento político-militar que derrubou o governo do presidente João Goulart, tem sido exaltado como *revolução* ou condenado como *golpe* de Estado, dependendo da fonte de informação. Isto poderia ser modelado como uma restrição `sem:View` sobre a propriedade `sem:eventType`, que atribui `sem:accordingTo` a uma `sem:Authority`. A classe `sem:Authority` é usada para indicar o 'de acordo com quem' dada assertiva é válida, e pode conter indivíduos da classe `sem:Actors` ou de outros vocabulários externos.

Apesar de fornecer meios para a modelagem de pontos de vista diferentes para eventos, SEM não consegue estabelecer relacionamentos entre eles, como causalidade, propósito ou correlação, e também não possui propriedades específicas que possam conferir intencionalidade ou agentividade (ativo/passivo) às participações dos atores. Nesse sentido, o

modelo SEM aproxima-se mais das ontologias Event e LODE do que das ontologias CIDOC e DUL, que prevêem um escopo mais amplo e rico para a instanciação de propriedades e relacionamentos em eventos.

4.8 Linking Open Descriptions of Events (LODE)

LODE[74] é uma ontologia simples e enxuta, desenvolvido originalmente para atender aos domínios da história e do jornalismo na modelagem dos seus aspectos "factuais", aspectos estes identificados pelas questões básicas: *what, where, when, who*, ou seja, o que aconteceu, onde aconteceu, quando aconteceu, quem estava envolvido (SHAW et al, 2009).

LODE contém uma classe, `lode:Event`, e seis propriedades: `atTime`, `circa`, `inSpace`, `atPlace`, `involved` e `involvedAgent`. Tanto a classe quanto as propriedades são mapeadas para outras ontologias de evento, como o CIDOC, EO e DUL, através do uso de `owl:sameAs` e `rdfs:subPropertyOf`. Assim LODE garante a interoperabilidade e o reuso de vocabulários externos mais complexos, enquanto mantém um número mínimo e enxuto de classes e propriedades.

Termo	Tipo	Definição
`Event`	classe	Algo que aconteceu, conforme relatado em um artigo de imprensa ou noticiado por um historiador.
`atPlace`	propriedade	Um lugar nomeado ou especificado relativamente, que é onde um evento aconteceu.
`atTime`	propriedade	Um instante abstrato ou intervalo de tempo, que é quando um evento aconteceu.
`circa`	propriedade	Um intervalo de tempo que pode ser precisamente descrito usando datas de calendário ou horas de relógio.
`illustrate`	propriedade	Um evento ilustrado através de alguma coisa (em geral, um objeto de mídia)
`inSpace`	propriedade	Uma região abstrata do espaço (um ponto ou região geoespacial) que é onde um evento aconteceu.
`involved`	propriedade	Um objeto (físico, social ou mental) envolvido em um evento.
`involvedAgent`	propriedade	Um agente envolvido no evento.

Tabela 4 – Ontologia LODE - Linking Open Descriptions of Events (LODE Documentation)

Pontos-de-vista podem ser expressos via mapeamento com o DUL através das entidades que definem situações (`dul:Situation`) e descrições (`dul:Description`). Tempo e intervalos de tempo são referenciados via mapeamento com a classe `time:TemporalEntity`. Tanto a propriedade `atTime` quanto `inSpace` são propriedades

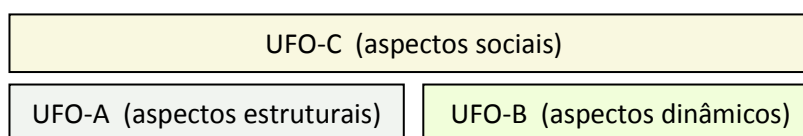
[74] http://linkedevents.org/ontology/

funcionais em OWL (`owl:FunctionalProperty`) significando que um evento só pode ser associado a no máximo um intervalo de tempo e um único lugar de ocorrência (VAN HAGE et al, 2011). Para modelar a participação de agentes ou objetos, LODE fornece as propriedades `involvedAgent` e `involved`, respectivamente, não considerando o caráter do agente (passivo/ativo) nem entidades abstratas para objetos. Além disso, foram excluídas propriedades que relacionam eventos pelas suas dimensões de causalidade, influência, propósito, ou ainda de mereologia.

4.9 Unified Foundational Ontology (UFO)

Em sua tese de doutorado, Guizzardi (2005) apresenta a proposta para uma ontologia de fundamentação fortemente apoiada sobre teorias das áreas de Ontologias Formais, Lógica Filosófica, Ciência Cognitiva e Lingüística, para dar suporte à modelagem conceitual, de uma maneira geral. Denominada Unified Foundational Ontology (UFO), hoje um projeto bastante maduro, ela tenta suprir limitações encontradas em outras ontologias de fundamentação, como a GFO e a DOLCE, unificando o que têm de positivo e acrescentando aspectos mais ricos ao esquema (GUIZZARDI; WAGNER, 2005), principalmente no que se refere à representação de entidades sociais. A ontologia foi dividida em três fragmentos: (i) UFO-A, para modelagem de objetos (endurantes) e suas propriedades; (ii) UFO-B, para eventos (perdurantes); e (iii) UFO-C, para sistematização de conceitos sociais e intencionais.

UFO-C (aspectos sociais)

UFO-A (aspectos estruturais) | UFO-B (aspectos dinâmicos)

Segundo Guizzardi e Gerd (2010), a UFO-A sistematiza conceitos presentes em estruturas taxonômicas, relações todo-parte, propriedades intrínsecas, propriedades relacionais, entre outros. Já a UFO-B trata dos conceitos relativos a estados, processos, eventos e relações temporais, e a UFO-C, construída sobre os fragmentos anteriores, conceitos como plano, ação, objetivo, agente, intencionalidade, comprometimento e compromisso. Em particular, os fragmentos que tratam sobre eventos (UFO-B) e entidades sociais (UFO-C) são indiscutivelmente importantes para reflexão deste trabalho.

Da mesma forma que DOLCE e OntoClean, UFO distingue a noção de *indivíduos* (particulares) da de *universais* (tipos). Os primeiros como entidades únicas que existem na realidade, e os segundos como padrões de características que podem ser instanciados em um

número diferente de indivíduos. Além disso, *substanciais* denotam os indivíduos existencialmente independentes (como uma pessoa, uma casa, Getúlio Vargas, o Congresso) e *modo* a instanciação de uma propriedade. Um modo é um indivíduo que só pode existir em outros indivíduos, e podem ser de dois tipos: intrínsecos (dependentes de um único indivíduo), por exemplo, uma cor, uma dor, uma temperatura; e relacionais (dependentes de vários indivíduos), por exemplo, um casamento, um tratamento médico, um casamento.

Eventos, na UFO-B, são indivíduos que se estendem no tempo acumulando partes temporais e também são entidades ontologicamente dependentes, ou seja, dependem de seus participantes para existir. Eventos podem mudar o mundo real ao mudar o estado das coisas de uma situação para outra; por exemplo, seja o evento *e: o ataque de Brutus a César*. Esse evento *e* é composto da participação individual das entidades César, Brutus e a faca usada no ataque. Cada uma dessas participações é por si própria um evento que pode ser atômico (não possui sub-partes) ou complexo (constituído de outros eventos), mas que para existir depende de um substancial (GUIZZARDI, FALBO & GUIZZARDI, 2008).

A figura a seguir mostra como os eventos podem ser analisados: i) como entidades que se estendem no tempo com certas estruturas mereológicas (eventos simples ou complexos); ii) como entidades ontologicamente dependentes que podem envolver um número de participações individuais.

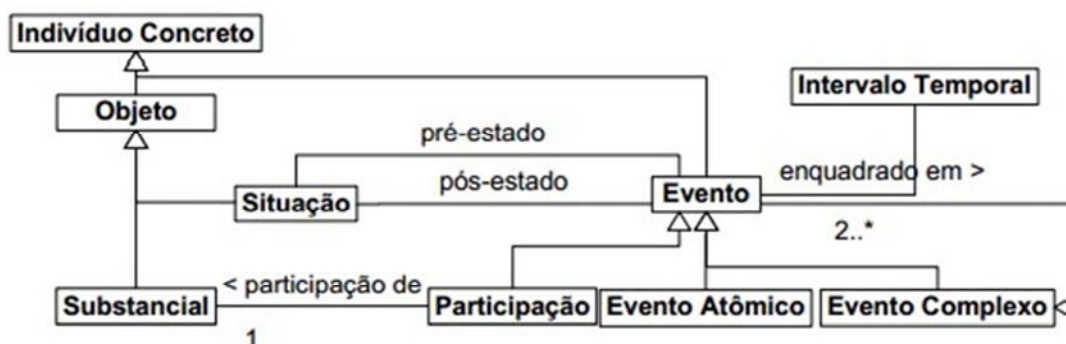



Figura 38 - Fragmento da UFO-B (Fonte: Guizzardi, Falbo & Guizzardi, 2008)

Quanto à localização no espaço, todos os modos espaciais de eventos são definidos em termos dos modos espaciais de seus participantes, e em contraste, todos os modos temporais de substanciais são definidos em termos dos eventos nos quais eles participam.

O modelo permite uma diversidade de estruturas temporais, tais como tempo linear, ramificado, paralelo e circular. Para o caso de estruturas ordenadas, a UFO se baseia em Allen (1983 apud Martins et al, 2011) como mostrado na figura 33:

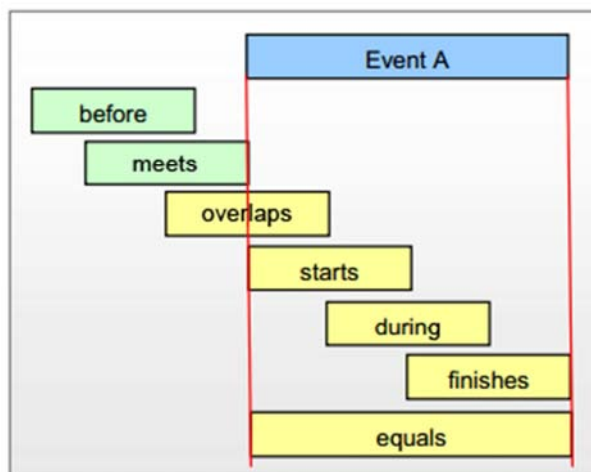



Figura 39 - Relações temporais entre eventos (Fonte: Allen, 1983 apud Martins et al, 2011)

As possíveis relações temporais nessa estrutura são: precede (*before*), encontra (*meets*), sobrepõe (*overlaps*), inicia (*starts*), durante (*during*), termina (*finishes*) e equivale (*equals*).

A UFO-C é construída sobre a UFO-A e UFO-B, e vem complementar o processo de representação de uma dada realidade, tratando aspectos dificilmente encontrados em outras ontologias, como desejo, crença, intenção. Para chegar a este modelo de abstração, os autores elaboram primeiramente uma descrição detalhada para situar os agentes e os relacionamentos possíveis em ocorrências de ação. Começa por distinguir substanciais agentivos (ou simplesmente, agentes) dos substanciais inanimados (não agentivos). Uma pessoa é um exemplo de agente físico, já uma organização seria um agente social; um livro e um carro seriam indivíduos substanciais inanimados; dinheiro e normas seriam substanciais inanimados sociais. Uma descrição normativa é um tipo de substancial inanimado social que define uma ou mais regras/normas reconhecidas por, pelo menos, um agente social. Um bom exemplo de descrição normativa seria a Constituição Brasileira.

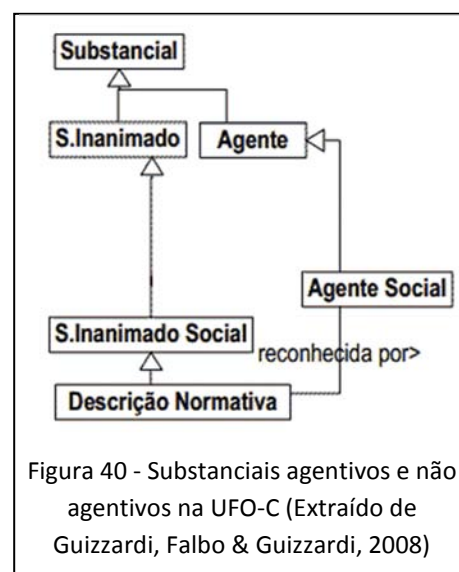



Figura 40 - Substanciais agentivos e não agentivos na UFO-C (Extraído de Guizzardi, Falbo & Guizzardi, 2008)

Escrever um artigo, realizar um negócio ou um ato comunicativo são exemplos de ações executadas por agentes. Elas podem ser atômicas ou complexas, sendo uma ação complexa composta de duas ou mais participações. Essas participações, por sua vez, podem ser intencionais ou não: por exemplo, o ataque de Brutus a César inclui a participação intencional do agente Brutus - que se pode denominar contribuição de ação - e a participação não intencional do substancial inanimado faca - que se pode denominar recurso. Dentre os tipos de participação de um recurso estão: i) criação, o objeto é criado após a ação; ii) término, o objeto deixa de existir após a ação; iii) alteração, o objeto é modificado durante a ação; iv) uso, o objeto participa da ação, sem que haja eliminação, alteração ou término.

No caso de uma ação complexa composta de contribuições de diferentes agentes, existe aí uma interação.

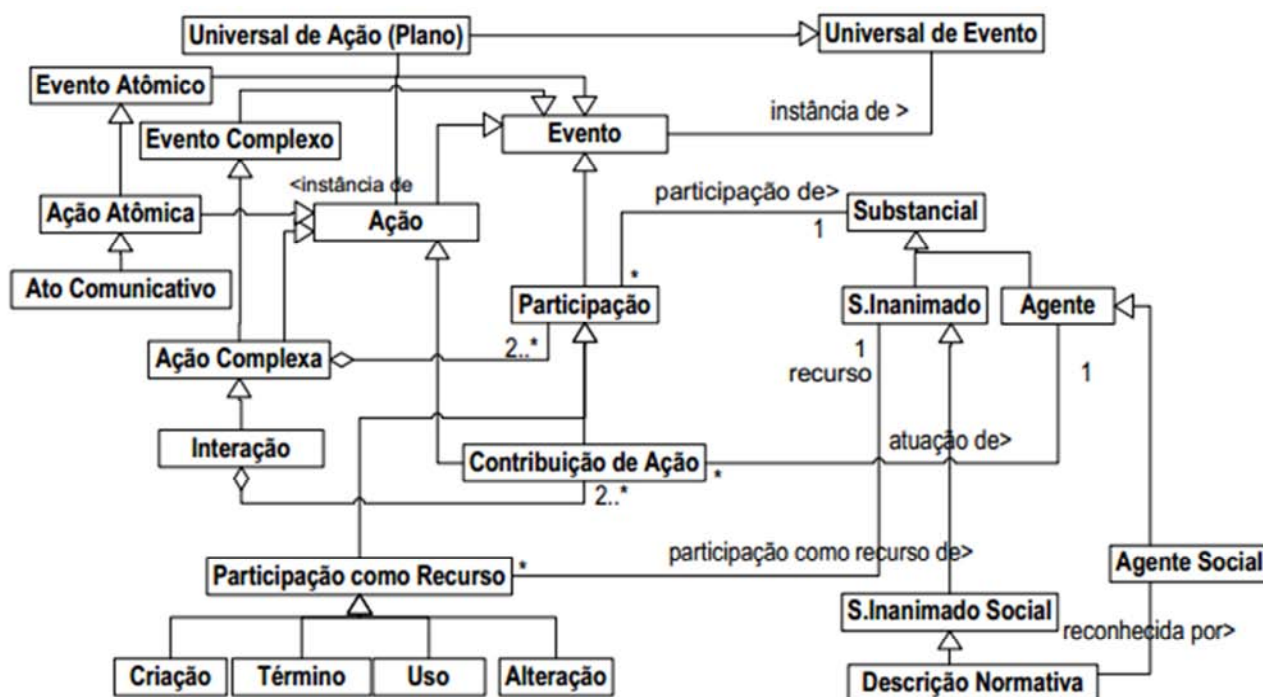



Figura 41 - Fragmento da UFO-C: Ações, Agentes e Substanciais Inanimados
(Extraído de Guizzardi, Falbo & Guizzardi, 2008)

Ações são eventos intencionais executadas por agentes, e estes, por sua vez, trazem em si modos intencionais que podem ser do tipo mental (crença, desejo, intenção) ou social (reivindicação, comprometimento social). Todo modo intencional contém uma proposição (conteúdo proposicional) cujo objetivo é satisfazer algum compromisso interno ou social do agente.

Figura 42 - Fragmento da UFO-C: Modos Mentais e Sociais
(Extraído de Guizzardi, Falbo & Guizzardi, 2008)

A combinação dos três fragmentos da UFO resulta em um modelo de representação bastante rico e flexível capaz de contemplar com riqueza semântica os conceitos presentes em uma dada porção da realidade, do nível mais concreto (indivíduos, particulares) ao mais abstrato (modos mentais, universais). Apesar da ontologia HIST não prever o tratamento de primitivas existentes no campo da ideologia ou intenção, é certo que em algum momento isso será revisto, dada as especificidades do domínio. Dessa forma, apesar de ser uma iniciativa relativamente recente que vem sendo aprimorado ao longo dos últimos anos, a UFO fornece contribuições importantes no processo de modelagem dos aspectos dinâmicos presentes no domínio da história.

4.10 VICODI Ontology

VICODI é o acrônimo de *Visual Contextualization of Digital Content*. O objetivo maior deste projeto é criar um ambiente de visualização baseado em tecnologias semânticas para busca, recuperação e exibição de informação, amparadas por contextos (NAGYPÁL, 2005a). O sistema tenta prever, baseado em ações ou escolhas prévias, aquilo que o usuário desejará ver na próxima tela. Conforme navega pelos resultados e realiza buscas, o contexto do documento corrente é usado para filtrar, selecionar e ranquear os próximos resultados e assim por diante. O uso de ontologia é o ponto-chave para a contextualização do domínio. Permitirá ao sistema armazenar conhecimento histórico de forma que seja possível aos algoritmos inferir e extrair informações discriminadas para o usuário.

Foram eleitas sete entidades, denominadas *Flavours*, para constituírem o núcleo da ontologia: `vicodi:Individual`, `vicodi:Object`, `vicodi:SocialGroup`, `vicodi:Organisation`, `vicodi:Event`, `vicodi:Location`, `vicodi:AbstractNotion`. A título de esclarecimento, noção abstrata abrange conceitos intangíveis, como ideias, ideologias, formas artísticas e construtos intelectuais.

Todos estes *flavours* têm associados a si uma instância da classe `vicodi:Papel`, que por sua vez é existencialmente dependente de uma instância de tempo (`vicodi:Tempo`). Intervalos de Tempo são representados por instâncias da classe `vicodi:TimeInterval`, que é uma subclasse de `vicodi:Time`.

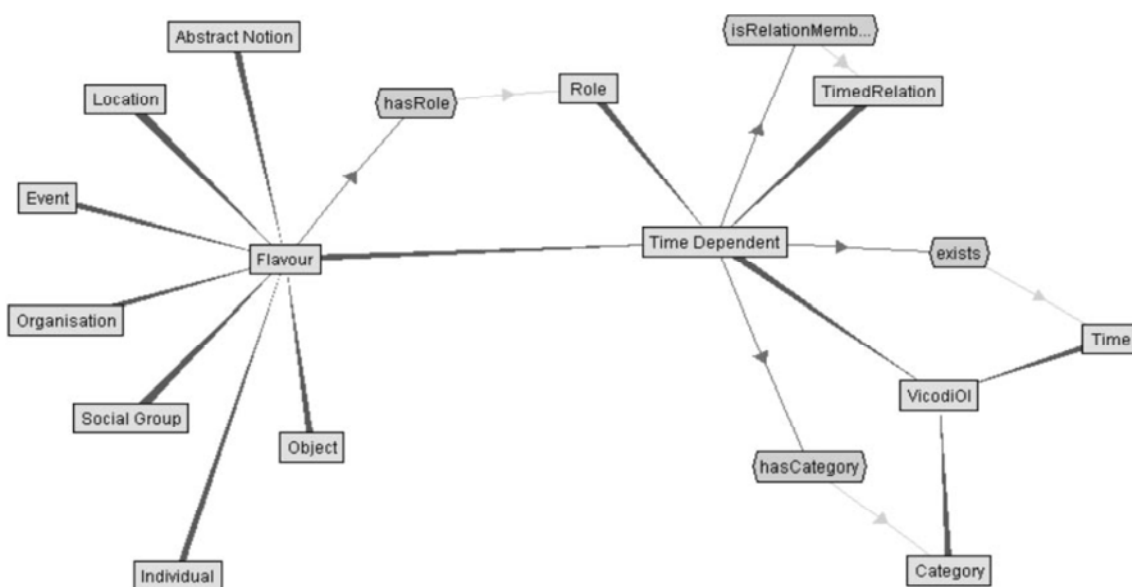



Figura 43 - A estrutura de primeiro nível da ontologia VICODI (Fonte:VICODI Documentation)

Para endereçar a questão da incerteza sobre quando certos eventos aconteceram, o modelo adota lógica probabilística e o conceito de fuzzyness[75]. A ideia básica é atribuir um peso de 0 a 1 principalmente às instâncias de dimensão local e temporal, e assim representar formalmente que um evento, por exemplo, pode ter se dado entre o ano X e o ano Y, ou seja, há uma probabilidade 0.8 na escala fuzzy de que aconteceu naquela data. Dessa forma os algoritmos conseguem raciocinar matematicamente sobre intervalos aproximados e posicioná-los relativamente no subconjunto de possibilidades existentes para aquele determinado fato.

[75] *Fuzzy* em inglês significa “vago”, “imprecisão”, conceito usado em situações onde não é possível responder simplesmente “sim” ou “não”.

A ontologia resultante tem cerca de 170 conceitos, 15 propriedades e mais de 15.000 instâncias, e baseou-se na análise empírica dedutiva de um corpus de 2.000 documentos. Está disponível para download no site oficial do projeto: http://www.vicodi.org.

4.11 DAML Time Ontology (Time)

A DAML Time Ontology[76] foi desenvolvida por Jerry R. Hobbs com contribuiçoes de George Ferguson, James Allen, Richard Fikes, Pat Hayes, Drew McDermott, Ian Niles, Adam Pease, Austin Tate, Mabry Tyson, e Richard Waldinger. Estes autores são conhecidos pelas suas contribuições em modelagens relativas à dimensão tempo. A ontologia fornece um vocabulário para expressar relações topológicas entre instantes e intervalos, que inclui conceitos e propriedades para medidas de duração, representação de freqüência e unidades de tempo de calendário e relógio. Originalmente desenvolvida para descrever conteúdo temporal de páginas web e serviços online, atualmente é bastante adotada para outros fins e referenciada por inúmeras ontologias de domínio.

Apesar de não conter relacionamentos explícitos com classes de eventos, ela foi concebida para gerar conexões entre quaisquer conceitos de um domínio localizados ou estendidos na dimensão temporal, incluindo claro, eventos de qualquer natureza.

As classes e propriedades explicitadas abaixo são por si só reveladoras quanto à estrutura e simplicidade da Time Ontology, e são resumidas na tabela de propriedades, domínio e escopo.

- TemporalEntity
    - Instant
    - Interval
        - ProperInterval
            - DateTimeInterval
- DurationDescription
- DateTimeDescription
- DayOfWeek

Property Name	Domain	Range
before	TemporalEntity	TemporalEntity
after	TemporalEntity	TemporalEntity
hasBeginning	TemporalEntity	Instant
hasEnd	TemporalEntity	Instant
inside	Interval	Instant
intervalEquals	ProperInterval	ProperInterval

[76] OWL Web Ontology Language 1.0 Reference. W3C Working Draft. http://www.w3.org/TR/owl-ref/

intervalBefore	ProperInterval	ProperInterval
intervalMeets	ProperInterval	ProperInterval
intervalOverlaps	ProperInterval	ProperInterval
intervalStarts	ProperInterval	ProperInterval
intervalDuring	ProperInterval	ProperInterval
intervalFinishes	ProperInterval	ProperInterval
intervalAfter	ProperInterval	ProperInterval
intervalMetBy	ProperInterval	ProperInterval
intervalOverlappedBy	ProperInterval	ProperInterval
intervalStartedBy	ProperInterval	ProperInterval
intervalContains	ProperInterval	ProperInterval
intervalFinishedBy	ProperInterval	ProperInterval
years	DurationDescription	xsd;decimal
months	DurationDescription	xsd;decimal
weeks	DurationDescription	xsd;decimal
days	DurationDescription	xsd;decimal
hours	DurationDescription	xsd;decimal
minutes	DurationDescription	xsd;decimal
seconds	DurationDescription	xsd;decimal
hasDurationDescription	TemporalEntity	DurationDescription
year	DateTimeDescription	xsd;gYear
month	DateTimeDescription	xsd;gMonth
week	DateTimeDescription	xsd;nonNegativeInteger
day	DateTimeDescription	xsd;gDay
dayOfWeek	DateTimeDescription	DayOfWeek
dayOfYear	DateTimeDescription	xsd;nonNegativeInteger
hour	DateTimeDescription	xsd;nonNegativeInteger
minute	DateTimeDescription	xsd;nonNegativeInteger
second	DateTimeDescription	xsd;decimal
timeZone	DateTimeDescription	tzont;TimeZone
inDateTime	Instant	DateTimeDescription
inXSDDateTime	Instant	xsd;dateTime
hasDateTimeDescription	DateTimeInterval	DateTimeDescription

Tabela 5 - Propriedades da DAML Time Ontology

4.12 Algumas considerações

Inúmeros modelos e ontologias que versam sobre eventos estão disponíveis atualmente. Eles diferem no foco, especificidade de domínio, complexidade e nível de formalização, e é possível criar mapeamentos para reuso de suas classes e propriedades. Um fato que se destacou nessa análise, aliás, é a confirmação de que todas elas procuram ajustar-se ao que já existe, ou seja, o reuso de ontologias é recorrente, mesmo se tratando de ontologias de topo. O trabalho ainda está em progressão para a maioria delas, pois o desenvolvimento de ontologias é orgânico e pressupõe esforço contínuo e de longo prazo. Percebe-se fortemente que a realização de escolhas que sejam mais adequadas ao modelo de representação de eventos não é tarefa simples, e o exercício de instanciação é trabalhoso e sujeito a equívocos, e por isso, a importância de se terem compromissos ontológicos bem estabelecidos.

5 Proposta de modelagem conceitual da ontologia

Tendo visitado os aportes teóricos e metodológicos recorrentes na área da representação do conhecimento nas suas variadas imersões disciplinares, neste capítulo conduziremos um estudo para a modelagem conceitual de um esquema inicial de ontologia no domínio da História. Apesar de os modelos nos quais este trabalho se apoiou estarem disponíveis em inglês, optou-se pela língua portuguesa para a apresentação das entidades propostas. Para fins de implementação no futuro, que vise seu reuso e interoperabilidade, certamente esta questão será reconsiderada.

Os mapas conceituais presentes no marco empírico foram elaborados com a ajuda do software Cmap Tools[77], e são formas usualmente adotadas para representação gráfica de modelos de conhecimento.

5.1 Fase de Especificação

Nesta etapa procuramos identificar o propósito, escopo e questões de competência da ontologia, conforme orientações fornecidas pelas metodologias levantadas. Em seguida, realizamos uma breve análise sobre os conceitos explicitados por estas questões.

1) Qual é o domínio a ser coberto pela ontologia?

   A História, e mais especificamente a História do Brasil contemporâneo.

2) Que propósitos a ontologia pretende atender?

   a) Normalizar vocabulário e facilitar a comunicação e o entendimento entre gestores e usuários de sistemas de informação;

   b) Permitir descrições mais ricas nestes sistemas;

   c) Constituir uma base de conhecimento que possa, após sua formalização, atender a portais semânticos e conectar-se ao Linked Open Data;

   d) Permitir seu reuso e integração com outras ontologias.

3) A quem se destina a ontologia? Quais usuários estarão envolvidos em seu uso?

[77] http://cmap.ihmc.us/

No caso das instituições que disponibilizam informação histórica, os usuários se dividem em dois grupos: os que vêm em busca de material de pesquisa e os que trabalham na sua organização e difusão. Os primeiros se constituem de pesquisadores, professores, alunos e leigos interessados na História do Brasil; e os últimos, de arquivistas e especialistas responsáveis pela atribuição de descrição dos recursos armazenados nos sistemas. No processo de construção, além dos especialistas da área, também é esperado o envolvimento de profissionais de tecnologia e ciência da informação.

4) Quais tipos de questões as informações contidas na ontologia deverão responder?

Para esta pergunta optamos por formular algumas questões de competência derivadas de cenários de motivação (GRUNINGER e FOX, 1995; NOY e MCGUINESS, 2001):

Cenários:

a) Deseja-se recuperar informações sobre determinado político, tais como, eventos nos quais tenha participado, documentos que produziu, papéis desempenhados nas organizações a que pertenceu, tipologia dessas organizações.

b) Quais eventos precederam ou sucederam certo acontecimento, como por exemplo, o assassinato de um político ou a deposição de um governo? Quais políticos da esfera militar estavam diretamente envolvidos e exercendo que papéis?

c) Deseja-se obter um perfil das pessoas que nasceram em dado período e ocuparam algum cargo político no Executivo. As de filiação partidária notadamente de esquerda frequentaram os mesmos círculos? As mesmas instituições acadêmicas? Tiveram a mesma formação? As mesmas influências?

d) Ainda sobre a elite política do país, como é possível descrever as características dos nossos representantes a partir dos seus dados sócio-econômicos e moradia? Quantos nasceram em áreas de carência (conforme dados do Censo) e quantos se estabeleceram em estados diferentes daqueles onde nasceu?

e) Deseja-se conhecer mais sobre a história de determinado local, enumerando cronologicamente os eventos que tiveram lugar, os políticos que ali nasceram e ocuparam cargos na esfera do Legislativo, e destes, quais compuseram a bancada de oposição ao governo e produziram algum documento relevante.

Algumas questões de competência:

Questões de competência	Identificação dos termos envolvidos
Quem é a pessoa P?	Pessoa, verbo Ser
Em qual(is) evento(s) a pessoa P se envolveu?	Evento, Pessoa, verbo Envolver
Qual o papel da pessoa P no evento E?	Papel, Pessoa, Evento
Em que tempo se deu o evento E?	Evento, Tempo
Qual a relação entre a pessoa P1 e a pessoa P2?	Pessoa, Relação
Que ação(ões) a pessoa P praticou e quando?	Ação, Pessoa, Tempo, verbo Praticar
O evento E1 conecta-se a que outro evento E2?	Evento, verbo Conectar
O que foi o evento E?	Evento, verbo Ser
Que pessoa(s) se envolveu(eram) no evento E?	Pessoa, Evento, verbo Envolver
Em que lugar ocorreu o evento E?	Evento, Lugar, verbo Ocorrer
Quando aconteceu o evento E? Quanto tempo durou?	Evento, Tempo, verbo Acontecer
O evento E1 tem relação com que outros eventos?	Evento, Relação, verbo Ter
Qual a relação existente entre o evento E1 e o evento E2?	Evento, Relação, verbo Existir
Qual(is) evento(s) ocorreu(eram) no local L e no tempo T?	Evento, Local, Tempo, verbo Ocorrer
Qual(is) pessoa(s) possui(em) relação com o local L?	Pessoa, Local, Relação, verbo Possuir
Que papél(is) a pessoa P exerceu no interior da instituição I?	Pessoa, Papel, Instituição, verbo Exercer
Quando a pessoa P ocupou o cargo C?	Pessoa, Cargo, Tempo, verbo Ocupar
Que cargo a pessoa P ocupava no momento do evento E? Que outras pessoas ocupavam o mesmo cargo?	Pessoa, Cargo, Evento, verbo Ocupar
Que pessoa(s) pertencia(m) à organização O?	Pessoa, Organização, verbo Pertencer
Em que lugar a pessoa P estudou?	Pessoa, verbo Estudar
Sobre o que trata o documento D?	Documento, verbo Tratar
Quem criou o documento D? Quando?	Pessoa, Documento, Tempo, verbo Criar
O documento D tem relação com que evento(s)?	Documento, Relação, Evento, verbo Ter
O processo PO durou quanto tempo?	Processo, Tempo, verbo Durar
Qual(is) pessoa(s) participaram do processo PO?	Pessoa, Processo, verbo Participar

Tabela 6 - Exemplos de questões de competência e termos envolvidos

5.2 Análise do conhecimento do domínio e proposição das classes principais

Dos termos derivados das questões de competência é possível distinguir aqueles que são existencialmente independentes daqueles que existem em função de outros (NOY & McGUINNESS, 2001): os primeiros constituirão as classes principais da ontologia e os segundos (em geral verbos e locuções verbais), acomodarão noções de propriedade para as instâncias dessas classes.

Dicionário de termos			
Termo	Definição geral	Tipo	Fontes de definição
Pessoa	Um indivíduo, alguém; personagem; ser real ou imaginário a que se atribui uma ação ou estado.	Conceito	FOAF, WordNet, Houaiss
Instituição	Estruturas parciais da sociedade que desempenham funções específicas na vida social. Formais ou informais, possuem certas normas que tendem a ser reconhecidas ou impostas socialmente.	Conceito	BIROU (1982)
Organização	Constituição de uma instituição pública ou particular; entidade com finalidades políticas, culturais, artísticas, etc.; estrutura organizada de pessoas com o propósito de administrar algo.	Conceito	FOAF, WordNet, Michaelis
Cargo	Incumbência; responsabilidade; função; posição formal no interior de uma organização.	Conceito	Michaelis
Papel desempenhado	Atribuição ou função *assinalada a*, ou *esperada de*, ou *desempenhada por* uma pessoa ou um grupo.	Conceito	WordNet, Michaelis
Local/Lugar	Área geográfica circunscrita e associada a algum tipo de estrutura política; local onde ocorre um evento; pode ser país, estado, cidade, etc.	Conceito	Michaelis, BIO
Documento	Declaração escrita que se reconhece como prova de um estado, condição ou acontecimento. Texto ou qualquer objeto que se colige como prova de autenticidade de um fato, ou que constitui elemento de informação.	Conceito	Houaiss, Michaelis
Tempo	Duração relativa das coisas; período contínuo no qual os eventos se sucedem; época (segundo os acontecimentos nele ocorridos).	Conceito	Houaiss, Michaelis
Evento	Algo fora do comum que ocorre em dado local e tempo, com certa duração; fenômeno localizado em um ponto singular de espaço-tempo; traz mudanças no estado de coisas e pessoas.	Conceito	WordNet, BIO

Processo	O curso de uma ação em particular com vistas a alcançar um resultado, um fim.	Conceito	WordNet
Ação	Processo dinâmico em que há um agente que faz algo; categoria que indica uma situação ativa do sujeito na proposição.	Conceito	Houaiss, WordNet
Relação	Tipo de vinculação de alguma ordem entre pessoas, fatos ou coisas; ligação, conexão, vínculo.	Conceito	Houaiss
(verbos)	Verbos são termos que exprimem uma ação, estado ou fenômeno da natureza, em geral usados para ligar o sujeito ao predicado.	Proprie-dade	Wikipedia, Michaelis

Tabela 7 - Dicionário de termos

Fundamentados na revisão de literatura e definição dos termos, podemos prosseguir com a análise:

- *Pessoa*, *Instituição* e *Organização* são entidades a quem se atribui uma ação, propósito ou propriedade: são agentes passivos e ativos em determinadas situações, e por essa característica de divisão podem ser definidos como renques de uma mesma superclasse. Que doravante denominaremos *Agente*.
- *Local* e *Lugar* são conceitos sinônimos na maioria dos dicionários. Em alguns contextos estes termos ganham significados distintos ('habitante local', 'variável local', 'lugar comum'), mas no geral têm a mesma denotação espacial, com a qual nos basearemos.
- Um *Documento* em si pode ser tratado de duas formas: como um objeto material, considerando seu aspecto físico e de suporte ou como um objeto imaterial, considerando seu conteúdo informacional. A ontologia deve ser capaz de modelar estas distinções, derivando-as de um conceito maior, conforme veremos adiante. Pela importância que *Documento* possui na História, ele constará como uma das classes principais.
- O conceito *Papel desempenhado* pode ser usado para designar a atribuição dada a alguém que possui determinado *Cargo*. Mas, além disso, papéis também instanciam pessoas (ou objetos) em outros contextos, como por exemplo, na participação de um evento ou em um relacionamento qualquer.

- *Evento*, *Ação* e *Processo* representam entidades que, de uma forma ou de outra - e em maior ou menor complexidade - existem ou persistem no tempo, e por isso também podem ser especializações de uma mesma classe.
- Uma *Relação* caracteriza a ligação entre indivíduos ou entre o indivíduo e um dado valor. Este termo 'carrega' uma propriedade, mas não é a propriedade em si no sentido de que não explicita o tipo de relacionamento existente; este se dará através dos *Verbos* das sentenças. Dessa forma, *Relação* é um conceito que se dilui em outros conceitos e os verbos denominam suas propriedades.

A partir destas definições derivam-se as primitivas para a nossa ontologia de História (HIST)**: Evento, Agente, Lugar, Tempo, Documento, Papel.**

Figura 44 - Primitivas da ontologia de História

Importante notar que esta categorização aproxima-se fortemente do modelo de Ranganathan (1967), expressas pelo acrônimo PMEST (personalidade, matéria, energia, espaço e tempo) conforme visto na seção 2.2.3.

5.3 Aquisição do conhecimento e conceitualização

No contexto deste trabalho a aquisição do conhecimento foi realizada de forma manual e em pequena escala face ao objetivo de ser apenas um protótipo-modelo conceitual de ontologia. Técnicas para extração automática de entidades nomeadas têm sido referenciadas em diversos trabalhos (BONTCHEVA & CUNNINGHAM, 2011), conforme já mencionado na seção 2.3.4.

Artigos, verbetes, listas de descritores e tabelas de bases de dados são fontes em potencial para a elicitação dos termos. O Centro de Pesquisa e Documentação de História

Contemporânea do Brasil da Fundação Getulio Vargas[78] (CPDOC/FGV) é um centro de referência dedicado ao estudo e à preservação da memória do país, detentora de um rico acervo neste sentido. Em especial, o Dicionário Histórico-Biográfico Brasileiro (DHBB), editado pela mesma instituição, apresenta-se como fonte preciosa, reunindo conhecimento acerca dos principais fatos e personagens que marcaram o Brasil pós-1930.

Para mostrar a forma como este processo pode ser orientado, alguns trechos que discorrem sobre os momentos que antecederam a morte do ex-presidente Getúlio Vargas servem como exemplos. Foram extraídos do DHBB e do livro "A Era Vargas", da cientista política Maria Celina D'Araujo:

> *Nos primeiros minutos do dia 5 de agosto, o jornalista Carlos Lacerda sofreu um atentado quando chegava à sua residência, na rua Toneleros, no Rio de Janeiro, em companhia de seu filho Sérgio e do major-aviador Rubens Florentino Vaz. O major Vaz, integrante de um grupo de oficiais da Aeronáutica que dava proteção ao jornalista durante a campanha eleitoral, teve morte instantânea, ao passo que Lacerda escapou com um ferimento no pé. O crime teve ampla repercussão no país, sendo imediatamente atribuído pela oposição a pessoas ligadas ao governo. Lacerda não hesitou em lançar imediatamente a culpa sobre o presidente. [...] Com a confirmação do envolvimento da guarda pessoal do presidente, a oposição intensificou sua campanha contra Getúlio (*CPDOC,2010).[79]

> *...segundo a imprensa, o governo passou a se ver mergulhado em um 'mar de lama'. Desgastado diante da opinião pública, Vargas escapa de uma nova deposição apelando para o suicídio. O tiro que desfechou no coração, no dia 24 de agosto de 1954, veio acompanhado de uma carta-testamento que se transformaria num dos mais conhecidos documentos históricos brasileiros* (D' ARAUJO,1997). [80]

Neles, percebemos a narrativa típica dos textos de história, onde cada fato aparece fortemente atrelado a outro, em uma estrutura de conexões que dá sentido ao todo. Analisando sintaticamente, é possível extrair e agrupar os conceitos que aparecem nos trechos pelas categorias levantadas anteriormente, e separar intuitivamente algumas das suas possíveis propriedades (nem todas necessariamente explicitadas acima):

[78] http://cpdoc.fgv.br. Trata-se da instituição onde a autora no momento atua profissionalmente.
[79] Trecho retirado do verbete Getulio Vargas, do Dicionário Histórico-Biográfico Brasileiro, e disponível em http://cpdoc.fgv.br/acervo/arquivospessoais/consulta.
[80] Trecho disponível em: http://cpdoc.fgv.br/producao/dossies/FatosImagens/GetulioVargas

Conceitos	Categoria	Atributos	Relacionamentos
Carlos Lacerda, Rubens Vaz, Getúlio Vargas, Grupo de oficiais da Aeronáutica, imprensa	Pessoa, instituição ou organização	tem nome, tem apelido, tem idade, tem gênero, tem data de nascimento, tem data de falecimento, tem data de fundação	participou de, é amigo de, é cônjuge de, morou em, formou-se em, estudou em, atuou como, é membro de, produziu, trabalha para, trabalha através de
Jornalista, major-aviador, oficial, presidente, segurança, oposição, pessoa ligada ao governo, guarda pessoal	Papel	tem descrição	é desempenhado por, é atribuído a, é subordinado a, tem subordinação
Atentado, crime, morte, campanha eleitoral, inquérito, deposição, suicídio	Evento, ação, processo	tem data de inicio, tem data de fim, tem descrição	tem participante, tem localização espacial, tem localização no tempo, sucede, precede, é associado a, é evidenciado por, é constituído por, tem natureza, tem finalidade
Carta-testamento, jornal (de imprensa)	Documento	tem data de criação, tem descrição, tem título	é sobre, tem autor, é referência de, é associado a, tem formato, é constituído de, é parte de, tem finalidade
5 de Agosto, 24 de agosto de 1954	Tempo	tem instante, tem descrição de tempo/data, tem descrição de duração	é localização temporal de, é constituído por, está contido em
Rua Toneleros, Rio de Janeiro, país	Lugar	tem coordenada	é localização geográfica de, é parte de, é constituído por, está contido em

Tabela 8 - Dicionário de conceitos

Os aspectos que parecem ser inerentes a todo evento histórico e devem ser tratados na ontologia são:

- aspecto constitutivo – o quê ou quem participa
- aspecto temporal – data, período, duração
- aspecto espacial – área geográfica, região, lugar relativo, localização
- aspecto estrutural – mereologia, causalidade, correlação
- aspecto evidencial – documentação produzida

5.3.1 Apresentação das classes principais

O esquema abaixo é uma representação das classes propostas para o domínio com alguns de seus possíveis relacionamentos. Na sequência, apresentamos uma breve definição de cada uma delas segundo sua inserção na história, uma análise de natureza ontológica, e ao final do capítulo, um quadro síntese com as propriedades levantadas.

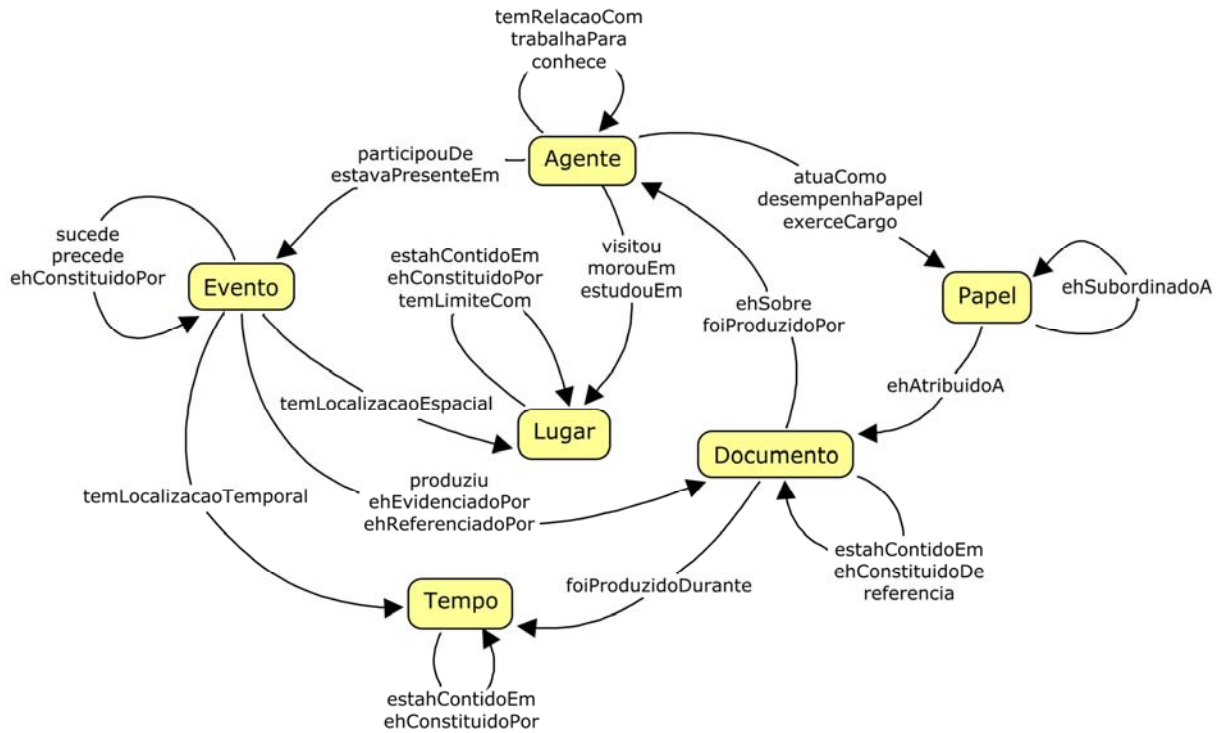



Figura 45 - Proposição de modelo de representação para o domínio da História

5.3.2 Classe: hist:Evento

Definição de Evento: a classe Evento descreve "o que" aconteceu. Um evento está contido em uma região espaço-temporal, que desejamos tratar como uma entidade a fim de fazermos asserções que o liguem a pessoas, lugares e documentos. Conforme já explicado anteriormente, consideramos evento todas as entidades que existem ou persistem no tempo, incluindo aí ação e processo. Apenas para fins de conceituação, especializamos a classe nos seguintes tipos emprestados da linguística: evento télico e processo.

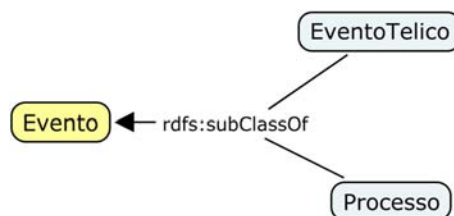



Figura 46 - Proposição da classe Evento

Evento télico se caracteriza por ter um ponto de culminação (um *telos*), isto é, um ponto de parada natural ou de conjunto conclusivo (CHIERCHIA, 2003). Na história seria o tipo que carrega de forma mais significativa a noção de evento em si, pois denota uma ocorrência marcante sob algum aspecto (BOX-SEFFENSMEIER e JONES, 2004; IDE e WOOLNER,

2007). Os eventos télicos mostrados abaixo foram obtidos com a ajuda das listas de descritores do acervo do CPDOC e do Dicionário Histórico-Biográfico Brasileiro, e representam termos recorrentes ao domínio.

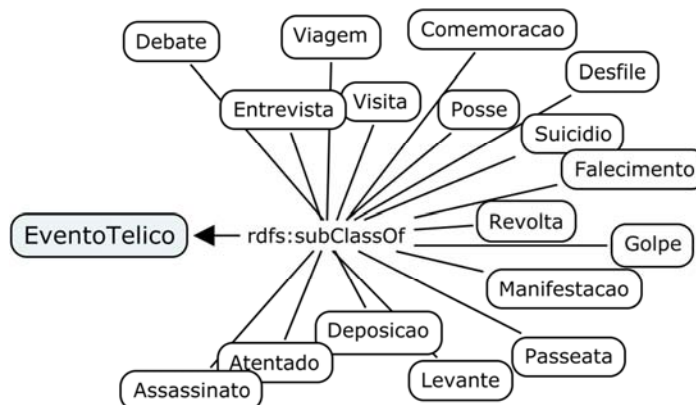



Figura 47 - Eventos de culminação

Processo denota uma sucessão sistemática de mudanças em uma direção definida (CHIERCHIA, 2003). Na história, um processo pode se constituir de um conjunto de fatos que apresentam certa unidade e evolui para um estado, positivo ou não, conclusivo ou não, sendo eminentemente políticos. Estão incluídas nesta categoria, além de metódos práticos adotados em sequência - como a censura, repressão, inquério policial militar, processo judicial, tortura, protesto e espionagem - linhas de pensamento ou ideologias que também implicam, de uma forma mais geral, em atividades contínuas ao longo do tempo, como o socialismo, o nacionalismo, comunismo e o anticomunismo.

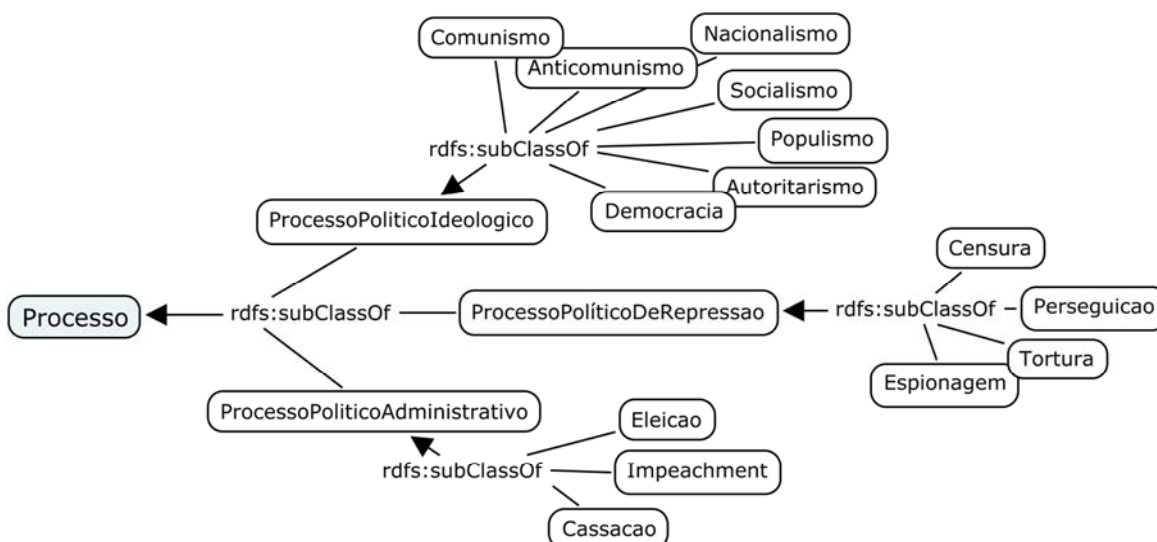



Figura 48 – Processos

Sobre a classe Evento:

Classes de referência de outras ontologias:	`event:Event; dul:Event; cidoc:E5.Event; edm:Event`
Atributos (`hist:`):	`temDataInicio, temDataFim, temDescricao, temNome`
Relacionamentos (`hist:`):	`temParticipante, temLocalizacaoEspacial, temLocalizacaoTemporal, sucede, precede, ehDeNatureza, ehAssociadoA, ehEvidenciadoPor, ehConstituidoPor`

Análise ontológica: Eventos são associados a uma data, período ou duração via `hist:temLocalizacaoTemporal`, e a lugares relativos (‘próximo a’, ‘dentro’, ‘longe’) ou referenciados (‘praça’, ‘Rio de Janeiro’, ‘país’) através da propriedade `hist:temLocalizacaoEspacial`. Eventos são relacionados aos agentes pela propriedade `hist:temParticipante` e a forma como se configura esta relação é especificada através do Papel desse agente. Cabe ressaltar aqui que o modelo é agnóstico no que se refere a julgamentos de aspecto, intencionalidade ou agentividade, sendo estes, quando existentes, sinalizados através dos documentos ligados aos eventos com as propriedades `hist:ehSobre`, `hist:ehReferenciaDe`. Quanto ao aspecto estrutural, ao contrário do modelo SEM, que distingue subeventos de um evento principal, consideramos a relação mereológica como uma relação entre eventos, cuja importância e partes constituintes serão explicitadas nas suas propriedades e relacionamentos. Nesse sentido, optou-se por seguir os modelos EO, DUL e CIDOC para que a classe receba instâncias de coisas que aconteceram em uma extensão limitada no tempo e que foram reportadas como sendo eventos, sejam eles de menor ou maior nível[81], conforme definição encontrada em (Cybulska e Vossen, 2011).

As estruturas apresentadas pelo OpenCYC e DOLCE comportam diferentes aspectos sobre eventos e procuram distingui-los de ações, processos e estados (situações estáticas), ao contrário de SEM e VIVO, que não explicitam esta separação em classes, mas permitem o uso de propriedades (por exemplo, `sem:eventType`) para instanciarem termos oriundos de vocabulários externos.

[81] Estes autores usam o massacre ocorrido na Sérvia em 1995, para distinguir eventos de menor e maior nível. A morte de uma criança no fogo cruzado ou a retirada à força dos homens para destino incerto, conforme noticiado na época, são eventos que se somam a outros para constituir o evento principal – Genocídio de Srebrenica –, descrito a partir de uma perspectiva de mais ‘alto’ nível que em geral só é possível obter no distanciamento fornecido pelo tempo.

Na UFO-B (GUIZZARDI e WAGNER, 2005), ações, eventos e estados são entidades perdurantes[82] que modificam o estado das coisas e devem ser tratadas a partir de noções temporais. Dos eventos instantâneos (atômicos) aos de longa duração, todos eles se estendem no tempo. Adotamos a mesma abordagem por também acreditar que as mudanças das coisas ou mudanças estruturais, que são a razão da história (BOX-SEFFENSMEIER e JONES, 2004), são viabilizadas tanto por uma ação quanto por um evento ou processo (SEWELL, 1996). Já um estado em si não tem a mesma denotação, mas ela pode ser apreendida a partir das conformações estabelecidas por várias das propriedades.

Allen e Ferguson (1994) explicam que uma mesma circunstância pode ser descrita por qualquer número de eventos. Por exemplo, a cena de uma bola que rola na mesa e vai ao chão pode ser descrita como 'a bola cai da mesa', ou 'a bola quica no chão', ou simplesmente 'a bola cai'. Nenhuma dessas descrições é mais correta que a outra, todas são modos de classificar o mundo e não propriedades inerentes a esse mundo (ALLEN & FERGUSON, 1994). Esta é a abordagem seguida pela DUL e posteriormente adotada pela LODE, que permite a uma dada situação referenciar vários eventos: `dul:includesEvent` → *domain*: `dul:Situation` e *range*: `dul:Event` (cardinalidade 1 para muitos). É uma forma interessante de tratamento, porém acreditamos que para atribuir interpretações, aspectos e pontos-de-vista distintos aos eventos, bastam as conexões entre eles e os recursos informacionais que os referenciam.

A natureza dos eventos tem sido investigada sob diferentes aspectos, e apesar de existirem definições isoladas para a Ciência da Computação, Inteligência Artificial e História, é possível observar uma característica em comum quanto ao mínimo que se deve esperar de um modelo de representação para eles: quem faz o quê, onde e quando.

5.3.3 Classe: hist:Tempo

Definição: Tempo é o tipo de dado que tenciona descrever "quando" algo aconteceu. Pode variar entre datas precisas (24 de agosto de 1954) ou incertas (década de 1950), a intervalos de tempo ou períodos (entre agosto e dezembro de 1954, quatro meses).

[82] Grosso modo, podemos definir perdurantes como sendo "processos" de um discurso; em contraponto, endurantes seriam os "objetos". Ao contrário dos endurantes (ex: uma pessoa, uma casa, a lua), os perdurantes compõem-se de partes temporais que podem ter diferentes propriedades (ex: uma tempestade, um ataque do coração, a I Guerra Mundial). (GUIZZARDI e WAGNER, 2005)

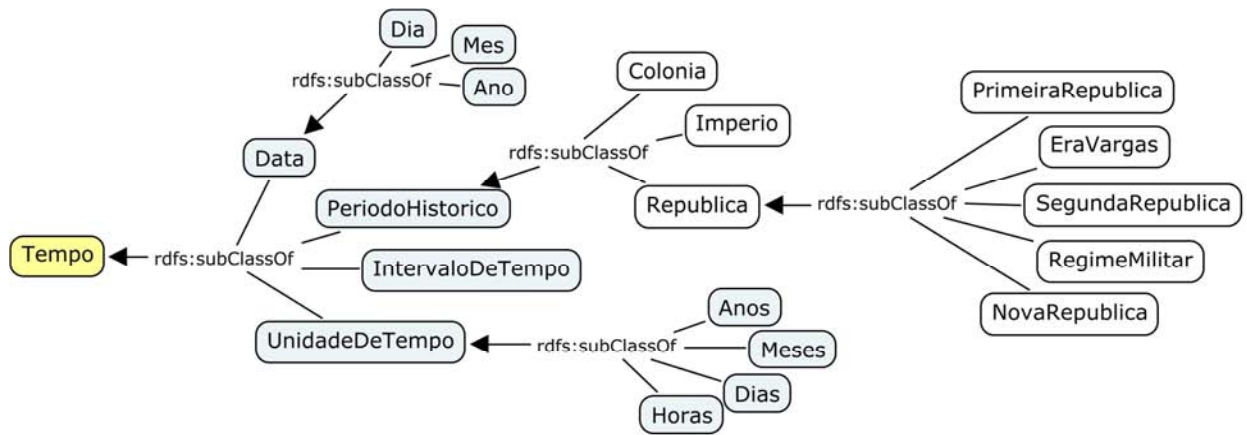



Figura 49 - Proposição da classe Tempo

Sobre a classe Tempo:

Classes de referência de outras ontologias:	`owltime:TemporalEntity; cidoc:E52.Time-Span; dolce:TimeInterval; edm:Time-Span; vicodi:Time`
Atributos (`hist:`):	`inicioDoIntervalo; fimDoIntervalo; duracaoDoIntervalo; data; hora; dia; mes; ano; horas; dias; semanas; meses; anos;`
Relacionamentos (`hist:`):	`ehLocalizacaoTemporalDe, ehConstituidoDe, estahContidoEm, ehReferenciadoPor, sucede, precede`

Análise ontológica: A história ‘preenche’ a dimensão tempo com indivíduos e suas atividades, e a análise do conhecimento do domínio sugere fortemente que esta dimensão seja representada também através de conceitos que expressem significado por algum tipo de associação, por exemplo, “Era Vargas”, “Regime Militar”, “Governo João Goulart”. São os denominados períodos históricos, que por sua vez também podem instanciar outras classes de Tempo (`hist:IntervaloDeTempo`, `hist:UnidadeDeTempo`).

Da mesma forma que a ontologia UFO se baseia nas relações temporais de Allen (1983 apud GUIZZARDI, FALBO & GUIZZARDI, 2008) para modelar estruturas ordenadas de eventos, também aqui estas relações estão presentes para representar intervalos (início/fim), instantes/datas (quando), concorrências ou simultaneidades (durante), coincidências (sobreposição), e sequências (antes/depois), permitindo o agrupamento ordenado de ações, eventos e processos.

As ontologias investigadas mostraram duas formas de inserir a noção de tempo em seus esquemas: uma, relacionando diretamente as instâncias de classes às propriedades *datatype*

usando literais RDF/XML; outra, definindo uma classe para representar intervalos temporais e usando propriedades de objeto para ligar as instâncias desta classe às instâncias das demais. A segunda forma é adotada pelo CIDOC, que introduz a classe `cidoc:E52.Time-Span` para representar intervalos temporais, e também pelo EO, que usa a classe `time:TemporalEntity` emprestada da W3C Time Ontology. Já a ontologia DUL permite a adoção tanto de uma quanto da outra forma: uma data pode ser atribuída a um evento utilizando diretamente o *datatype* `dul:hasEventDate` ou uma instância da classe `dul:TimeInterval` representando o dado período.

A vantagem de se associar datas diretamente aos eventos, segundo (SHAW, TRONCY, HARDMAN, 2009), tem a ver com simplicidade: é mais fácil filtrar, ordenar e usar rotinas de parseamento e comparação quando os dados seguem um padrão. Porém o problema surge quando temos noções mais complexas de tempo para representar, tais como períodos que não coincidem com unidades de data ou incerteza sobre quando tal evento ocorreu. Os autores da ontologia LODE estabeleceram que, onde possa existir desacordo sobre o intervalo de tempo associado a um evento, este desacordo deve ser modelado em um nível interpretativo para além do escopo da classe Tempo, e o valor de `lode:atTime` ou deverá ser especificado com o menor intervalo temporal que inclua as interpretações conflitantes, ou não ser especificado (SHAW, TRONCY, HARDMAN, 2009).

O modelo SEM define a entidade `sem:TimeStamped` como a superclasse de todas as outras classes – Evento, Pessoa, Objeto, Papel, etc – explicitando sua condição de dependência existencial na dimensão tempo. Diferentemente, e inspirado nas ontologias LODE e EO, o aspecto temporal de nosso modelo se estende às outras classes através da relação direta destas com a entidade `hist:Tempo`, apropriando instâncias de propriedades equivalentes às definidas na W3C Time Ontology: `temInicio`, `temFim`, `inicioDoIntervalo`, `fimDoIntervalo`, `duracaoDoIntervalo`, `descricaoDoIntervalo`, `data`, `hora`, `dia`, `mes`, `ano`, `horas`, `dias`, `semanas`, `meses`, `anos`.

Além da W3C DAML Time Ontology[83], existem muitas outras ontologias que da mesma forma desenvolveram meios para representar noções de tempo, como a SUMO[84], a

[83] http://www.w3.org/TR/owl-time/
[84] http://virtual.cvut.cz/kifb/en/concepts/_temporal_relation.html

TimeLine Ontology[85], Kestrel Time Ontology[86] e a Simple-Time Ontology[87]. Fernández-Lopez e Gomez-Pérez (2004) realizaram um levantamento do estado da arte sobre ontologias de tempo destacando características que podem ser úteis na análise de modelos temporais considerando reuso, granularidade e complexidade. Tendo em vista estes aspectos e existência de farta documentação, a W3C DAML Time Ontology foi a que os autores consideraram melhor candidata para o projeto de aplicação web que estavam desenvolvendo na época.

Sowa (2000) usa a dimensão tempo para formalizar, em termos axiomáticos, a natureza das entidades de um domínio, distinguindo continuantes e ocorrentes de forma semelhante aos endurantes e perdurantes da UFO e DOLCE. O autor recorre a (SIMONS, 1987 apud SOWA, 2000), que formula as seguintes definições: um continuante é um objeto que está no tempo, mas que não faz sentido dizer que possui partes temporais ou fases. Em qualquer tempo no qual ele exista, um continuante está totalmente presente. Um ser humano, por exemplo, está constantemente ganhando e perdendo moléculas, mas em qualquer tempo *t* quando uma pessoa *p* existe, todas as partes de *p* existem ao mesmo tempo *t*. Já os ocorrentes, que abrangem eventos, processos, acontecimentos e estados, localizam-se, como os continuantes, no tempo, mas ao contrário deles, possuem partes temporais. Concertos, eventos esportivos, viagens, tempestades, terremotos são exemplos de ocorrentes. Um ser humano é um continuante, no entanto a vida desse ser humano é um ocorrente cujos estágios estão espalhados entre o intervalo que vai do nascimento à morte.

Assim, a perspectiva de existência no tempo é fator que confere identidade aos conceitos do domínio. Também o aspecto mereológico é importante e deve ser considerado: se certo evento A é parte de um evento B, isso implica em que o período de tempo de B contém o período de A. Mas fazer parte é mais que uma contenção de tempo. Alguém pode se casar durante a Revolução de 1930, mas isso não faz desse casamento parte da Revolução. Assim, modelagem de eventos deve ser capaz de distinguir entre contenção meramente temporal e relacionamentos mereológicos entre subeventos e um evento maior.

O aspecto desafiador no domínio da história reside na adequada modelagem temporal dos conceitos a fim de derivar mudanças sobre eventos, ações e identidade.

---

[85] http://purl.org/NET/c4dm/timeline.owl#
[86] http://www.kestrel.edu/DAML/2000/12/TIME.daml
[87] http://www.ksl.stanford.edu/ontologies/time/

5.3.4 Classe: hist:Lugar

Definição: Em história, lugar é a classe que descreve "onde" algo aconteceu. Há diferentes formas de trabalhar a noção de espaço: como espaços delimitados política e/ou administrativamente (país, estado, município, bairro, logradouro), como lugares referentes (construção, monumento) ou naturais (rio, montanha), como coordenadas (latitude, longitude) e como regiões abstratas (perto, longe, dentro, em cima, embaixo).

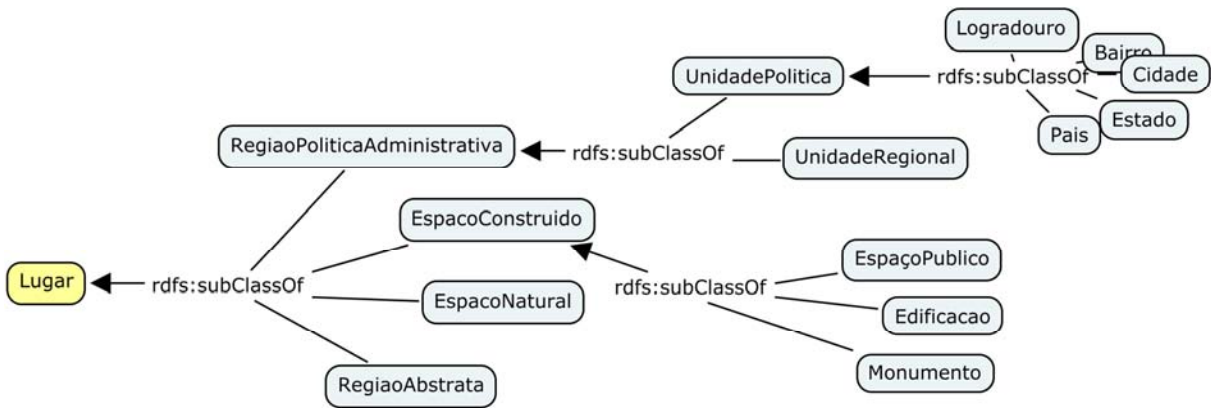



Figura 50 - Proposição da classe Lugar

Nossa abordagem prevê mais a forma de entidades administrativas (regiões, estados, sedes, etc) e menos a forma de entidades físicas (montanhas, oceanos, rios, etc) ou abstratas (dentro, fora, perto), pois é o tipo de informação espacial que tende a predominar no domínio. É raro encontrar eventos contemporâneos da história do Brasil, sobretudo os políticos, cujas referências de lugares não estejam contidas nesta forma de instanciação.

Sobre a classe Lugar:

Classes de referência de outras ontologias:	`dul:Place; cidoc:E53.Place; cidoc:E44.Place_Appellation; edm:Place`
Atributos (`hist:`):	`temCoordenada, temNome`
Relacionamentos (`hist:`):	`ehLocalizacaoEspacialDe, ehConstituidoPor, estahContidoEm, ehReferenciadoPor, temLimiteCom`

Análise ontológica: A ampla variedade de abordagens espaciais no domínio impactará diretamente o tipo de informação a ser revelado ou descoberto. Por exemplo: lugar, orientação e coordenadas permitem responder questões sobre direção ou distância de um lugar para outro. A maioria das ontologias analisadas permite fazer a distinção entre uma região espacial

abstrata e um lugar que tenha denominação/alcunha estabelecida. A DUL, por exemplo, dispõe de `dul:SpaceRegion` e `dul:Place`; já a ontologia CIDOC concebe as classes `cidoc:E53.Place`, `cidoc:E44.Place_Appellation` e as propriedades `cidoc:is_identified_by`, `cidoc:overlaps_with`, `cidoc:borders_with` e `cidoc:falls_within`. De um modo geral, parecem ser suficientes para delimitar e identificar bem o espaço onde se deu determinado evento.

'Podem os eventos se moverem?' Partindo desta provocação, Casati (2010) traça algumas considerações interessantes sobre a relação que liga evento a lugar. Ele argumenta que, apesar de eventos possuírem uma localização espacial da mesma forma que as entidades materiais também possuem, esta ocupação é qualificada de maneira distinta. Por exemplo, uma cadeira pode estar em determinado prédio num tempo em particular, mas um assalto pode ocorrer dentro do estacionamento, e eventualmente terminar num shopping. Como o autor considera todas as situações um único evento, ele pergunta: como tornar um só a soma dos *n* lugares onde ele se deu?

Como vimos anteriormente (GUIZZARDI e WAGNER, 2005; SOWA, 2000; MASOLO et al, 2002), eventos são classificados como entidades ocorrentes ou perdurantes, ou seja, eles se compõem de partes temporais podendo cada parte ter diferentes propriedades. Acreditamos que um evento possa se constituir de subeventos onde a cada um deles são atribuídos tempo, lugar e demais relacionamentos. Optamos por descartar, assim, a existência de múltipla cardinalidade para a propriedade `hist:temLocalizacaoEspacial` sob domínio (*domain*) `hist:Evento` e escopo (*range*) `hist:Lugar`. Um evento ocorre em um único lugar distinto.

5.3.5 Classe: hist:Agente

Definição: Agente descreve "quem" fez ou sofreu algo no contexto de algum evento ou processo histórico, podendo ser uma pessoa (personalidade), uma instituição ou organização.

A classe considera: i) Pessoas como indivíduos reais as quais se atribuem uma ação ou estado; ii) Instituições e organizações como estruturas formais ou informais da sociedade, podendo ou não ter uma base física. Ambas as entidades (instituição e organização) carregam definições que na prática se confundem entre si, até na instanciação particular de cada uma, e por isso o critério de usá-las conjuntamente, como sendo as entidades responsáveis pela organização das interações sociais e/ou estatais, servindo a ações de interesses políticos,

militares, econômicos, culturais, artísticos, internacionais. Como exemplos, podem-se citar o Serviço Nacional de Informação (SNI), o Departamento de Ordem Política e Social (DOPS), a Associação Brasileira de Imprensa (ABI), o Ministério do Exécito, o Superior Tribunal Militar, a Ordem dos Advogados do Brasil, os partidos políticos, os movimentos e as agremiações diversas.

É posível identificar entre as instituições e organizações, aquelas que são de caráter civil das que são de caráter estatal, ambas alicerçadas sob viéses tais como religiosos, políticos, militares, paramilitares e econômicos.

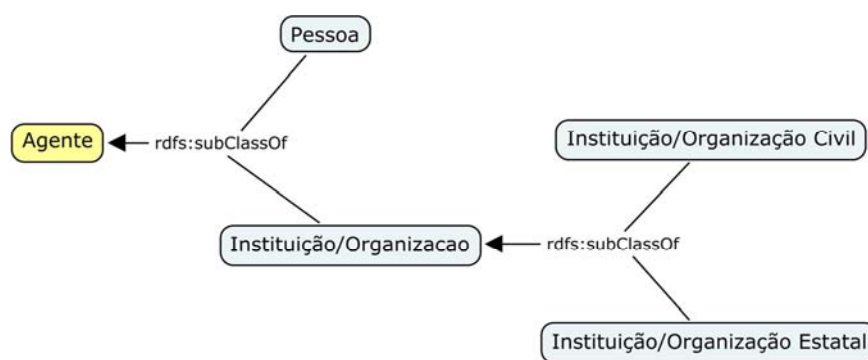



Figura 51 - Proposição da classe Agente

Há que se mencionar que a modelagem de entidades humanas, grupos e organizações, tem sido tratada com muita propriedade pelo projeto Friend-of-a-Friend[88]. FOAF é um vocabulário expresso através de RDF e OWL que busca descrever um agente, suas atividades e seus relacionamentos com outros agentes e objetos. O conjunto de assertivas, ou *fatos*, neste formato, dá origem à base de conhecimento sobre a qual é possível realizar inúmeras inferências automáticas:

- "Se Gregório Fortunato contratou capangas para assassinarem o jornalista Carlos Lacerda".
- "Se Carlos Lacerda era inimigo de Getúlio Vargas".
- "Se Gregório Fortunato trabalhava para Getúlio Vargas como chefe da sua guarda pessoal".

→ Então, é muito provável que Getúlio Vargas esteja relacionado com o atentado a Carlos Lacerda.

E ainda, conexões do tipo "encontre pessoas que estudaram na Faculdade de Filosofia, Ciências e Letras da Universidade de São Paulo entre 1934 e 1985 e exerceram algum cargo político no país entre 1964 e 1985".

Sobre a classe Agente:

---

[88] http://www.foaf-project.org/

Classes de referência de outras ontologias:	foaf:Agent; cidoc:E39.Actor; dul:Agent; edm:Agent;
Atributos (hist:):	temNome, temApelido, temIdade, temGenero, temDataDeNascimento, temDataDeFalecimento temDataDeFundacao, temDataDeExtincao
Relacionamentos (hist:):	temPapel, temCargo, trabalhaPara, ehAutorDe, ehReferenciadoPor, conceitua, ehConceituadoPor, encontrou, sofreu, atuouEm, participouDe, ehMembroDe, temMembro, atuaAtravesDe

Análise ontológica: Diretamente equivalente ao `foaf:Agent`, nossa classe `hist:Agente` também corresponde ao `sem:Actor` e `cyc:Agent-Generic`. Assim como ocorre nessas classes, a participação em um evento como agente ativo ou paciente é modelada via a classe Papel.

A propriedade `hist:temParticipante` liga os eventos aos agentes envolvidos. Ela é diretamente equivalente ao `dul:involvesAgent`, `cidoc:P11.had_participant` e `eo:agent`.

Com a propriedade `hist:atuaAtravesDe`, uma organização pode ser representada em diversas áreas distintas através das suas unidades de negócio e departamentos, e manter uma identidade única como Organização.

5.3.6 Classe: hist:Papel

Definição: Segundo alguns léxicos[89], papel representa uma atribuição ou função *assinalada a*, ou *esperada de*, ou *desempenhada por* uma pessoa ou um grupo. No domínio da história, papéis são de fundamental importância porque dão identidade às instâncias justificando sua existência como tais. Getúlio Vargas foi um advogado desde que se formou em Direito em 1907, um líder civil da Revolução de 1930, o presidente do Brasil em dois momentos – de 1930 a 1945 e de 1951 a 1954 - e um suicida em 1954. Como podemos ver, instâncias podem ter diferentes papéis de acordo com o contexto no qual estão inseridos, e sempre associados à dimensão temporal.

Assim, baseado na natureza dos contextos que determinam os papéis, consideramos a seguinte distinção básica (LOEBE, 2005; MASOLO et al, 2004): papéis processuais, que são

---

[89] Dicionário Michaelis, WordNet

formas de participação em um evento; e papéis sociais, legitimados ou construídos socialmente. Parecem-nos suficientes para os espaços de uso na história.

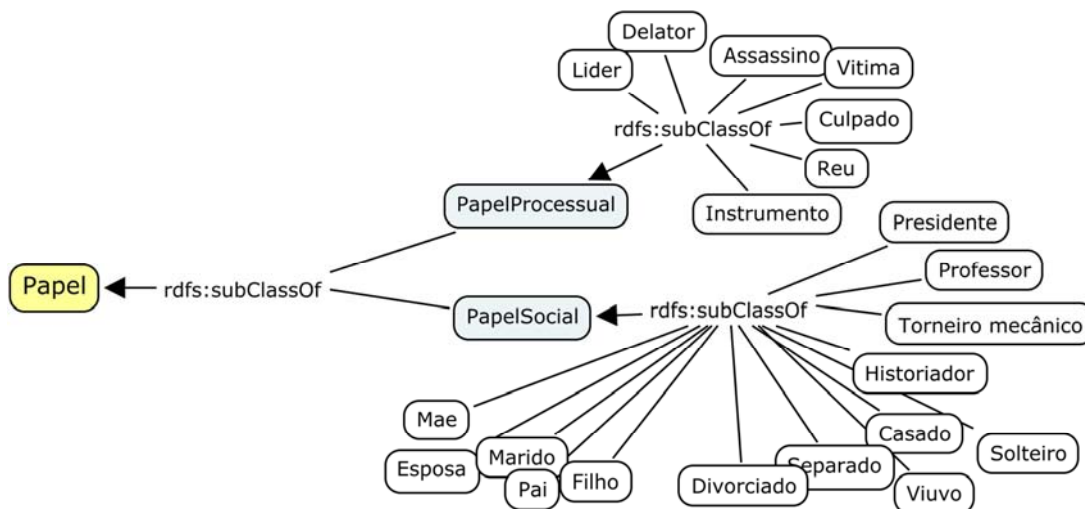



Figura 52 - Proposição da classe Papel

Sobre a classe Papel:

Classes de referência de outras ontologias:	`vicodi:Role, dul:Role; edm:Role`
Atributos (`hist:`):	`temDescricao`
Relacionamentos (`hist:`):	`ehPapelDe,ehReferenciadoPor, ehAtribuidoA, temSubordinacao, ehSubordinadoA`

Análise ontológica: Uma restrição proposta por Masolo et al (2004) é que papéis têm a característica de serem anti-rígidos (~R), ou seja, são propriedades contingenciais e não essenciais para todas as suas instâncias. Seguindo esta definição, J. Fan (2001, apud MASOLO et al, 2004) limita os papéis à “representação de características extrínsecas de uma entidade face ao seu envolvimento em um evento”, ou seja, papéis atribuem modalidades variadas de participação. Mas Loebe (2005) aponta que papéis não necessariamente precisam estar limitados ao tempo da participação em eventos específicos: um advogado continua sendo um advogado mesmo dormindo ou exercendo outras funções: seriam os papéis sociais.

Para reforçar, Masolo e Guarino (2002) explicam que os endurantes, dos quais fazem parte as pessoas e outros ‘objetos’ físicos, podem genuinamente mudar ao longo do tempo, ou seja, possuir propriedades incompatíveis em diferentes tempos. Quando uma instância possui qualidades específicas, ela as possui em determinada localização no tempo ainda que isto

signifique que seja durante toda a sua existência. Ou seja, a identidade do agente está intrinsecamente associada a instâncias do Tempo.

Guarino e Welty (2004) consideram papéis como propriedades, as quais podem ser atribuídas meta-propriedades. Como isso acontece? Na verdade os autores elaboraram formas de verificar a adequação ontológica dos relacionamentos entre entidades segundo noções extraídas da filosofia, como rigidez, identidade, unidade e essência. Estas noções, ou meta-propriedades, são usadas para caracterizar os aspectos relevantes do significado pretendido para as entidades, configurando-se em uma metodologia independente de domínio (OntoClean). Tome-se como exemplo a meta-propriedade rigidez: ela tem origem na noção filosófica de essência, podendo ser classificada como rígida (+R) e anti-rígida (~R). Uma propriedade é rígida se ela é essencial para todas as suas possíveis instâncias (por exemplo, todo humano é necessariamente um ser humano); é anti-rígida se não for essencial para todas suas instâncias (por ex, ser um sindicalista é anti-rígido porque a qualquer momento uma instância de Pessoa pode deixar de ser sindicalista para se tornar um presidente).

Segundo Sowa (2000), a identidade de uma entidade depende do conceito usado para caracterizá-la. Assim, para propósitos legais, um ser humano - que o autor classifica como sendo um continuante[90] -, pode ser considerado uma pessoa cuja identidade permanece constante por toda a vida. Mas para outros propósitos, a mesma pessoa pode ser classificada em diferentes tempos como um bebê, um adolescente, um adulto. Cada papel ressalta um aspecto diferente do que significa ser um ser humano. Cada um pode ser apropriado em algum contexto; mas dependendo da essência desse papel, se são aplicados ao mesmo indivíduo ao mesmo tempo, podem surgir contradições (SOWA, 2000). Por exemplo, ser ao mesmo tempo um bebê e um adulto (fisiologicamente falando) é impossível.

Na UFO-A, Guizzardi diferencia o que ele denomina como sortais anti-rígidos: fases (*phases)* e papéis (*roles)*. Fases são constituídas de partes temporais, caracterizando-se, como o próprio termo indica, uma fase passageira de um universal. Já papéis são processos, ou seja, funções executadas por entidades em determinados contextos ou por um período de tempo. Seria exemplo de fase a adolescência e de papel ‘ser empregado’ (GUIZZARDI, 2005).

---

[90] Um *continuante,* semelhantemente ao *endurante,* é um objeto que retém sua identidade ao longo de um extenso período de tempo; ao contrário, um *ocorrente* é um processo em constante mutação cujos estados evoluem de um para outro (SOWA, 2000).

Na lingüística, papéis são essencialmente estudados como papéis temáticos. Este seria o termo a designar o tipo de relação semântica que se encontra associada aos argumentos de um predicador, isto é, ao sujeito e aos complementos de um verbo, de um substantivo ou adjetivo. Em sentenças que descrevem situações ou eventos, Chierchia (2003) associa o conceito de papéis temáticos aos verbos, traduzindo-os como um conjunto de conseqüências determinadas pelas posições de uma relação: i) Maria observa Léo. ii) Maria beija Léo. iii) Léo abandona Maria. Em todas estas sentenças, é verdadeiro que Maria faz algo (agente) e algo é feito a Maria (paciente). Chierchia observa que há vários papéis temáticos e que a terminologia que os identifica pode variar. A que é proposta por ele, recuperando outros lingüistas, é a seguinte: Agente, Tema, Experienciador, Receptor, Origem, Instrumento, Beneficiário.

Sowa (2000) usa a clássica sentença "*Brutus stabbed Caesar violently with a shiny knife*" para enumerar as relações de agente (Brutus), paciente (Caeser) e instrumento (knife) correspondentes aos papéis temáticos usados pelos lingüistas. Nesse caso papéis também são imputados a objetos materiais.

Nosso modelo compreende as noções descritas acima para uso de atribuição de papéis às entidades do domínio.

5.3.7 Classe: hist:Documento

Definição: Documento é um objeto material (ou virtual, nos dias atuais) que carrega em si um conteúdo informacional. Na história o seu papel é de fundamental importância, podendo ser reconhecido como prova de uma condição ou acontecimento, além de constituir elemento de informação, pura e simplesmente.

Documento, da mesma forma que livro e obra, é um termo polissêmico que abrange pelo menos dois significados ou manifestações: artefato físico e entidade abstrata. O primeiro pode ser percebido através de características como formato, tamanho, número de páginas, etc, e identificado pela sua localização espaço-temporal. Afinal, dois volumes não podem ocupar o mesmo lugar ao mesmo tempo. Já a noção abstrata é independente do espaço e do tempo, sendo identificado pela autoria, título, data, função e outros critérios. Dessa forma, nenhuma instância pode pertencer a ambas as classes; elas são disjuntas. Quando falamos 'o documento está rasgado' claramente nos referimos ao objeto material ou artefato físico, mas se a sentença for 'li o documento', trata-se da entidade abstrata ou artefato mental.

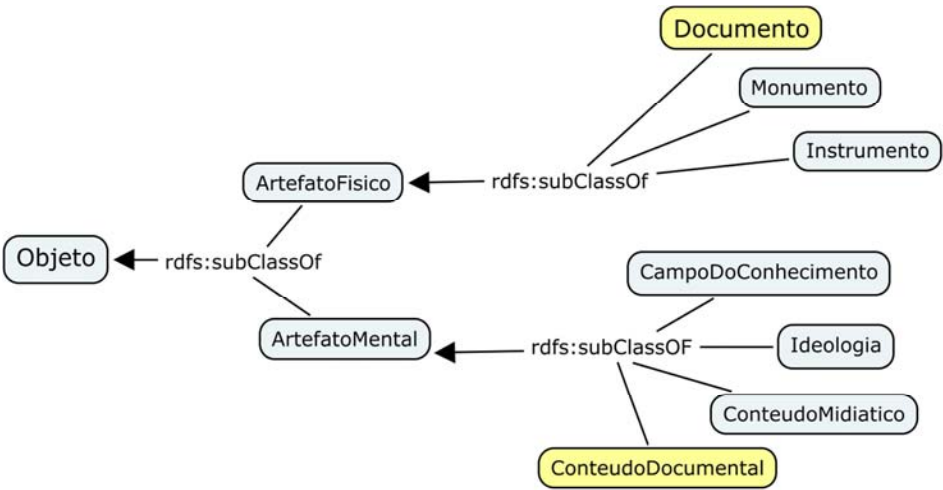



Figura 53 - Proposição da classe Documento

Sobre a classe Documento:

Classes de referência de outras ontologias:	cidoc:E31.Document; edm:InformationResource, vicodi:Writing
Atributos (hist:):	temDataDeCriacao, temDescricao, temTitulo
Relacionamentos (hist:):	temFormato, temAutor, ehSobre, estahContidoEm, ehConstituidoPor, ehAssociadoA, ehReferenciaDe, temReferencia, ehConsideradoComo, temFinalidade, foiProduzidoDurante, representa

Análise ontológica: A forma que adotamos para modelar as duas manifestações da entidade hist:Documento é especializando seu tipo hist:Objeto nas classes hist:ArtefatoFisico e hist:ArtefatoMental.

Documentos enquanto artefatos mentais viabilizam a transmissão de mensagens. Eles visam tanto à articulação e coordenação de atividades quanto ao registro de experiências e a utilização, reutilização e recontextualização do conhecimento. O processo é desencadeado por um produtor - muitas vezes denominado autor - segundo uma intencionalidade ou significado pretendido (MARCONDES, 2010). A Ciência da Informação estabelece lugar central para discussões sobre a gênese e contextos de uso do documento. São nas suas estruturas morfológica (significante) e semântica (significado) que estão constituídas as bases epistemológicas para a representação da informação (GONZALEZ, 2010).

Apoiados principalmente pela abordagem de viés antropológico que toma os documentos como fonte histórica, adotamos como premissa incluir no modelo apenas documentos de caráter eminentemente histórico, ou seja, aqueles realmente produzidos a partir de eventos históricos ou legados por algum agente atuante na história. O objetivo da ontologia não é explicar ou validar o conhecimento histórico em si (abordagem epistemológica), mas sim, como o próprio termo indica, estruturá-la como uma base (ontológica) de fatos e evidências de onde será possível inferir conhecimento do domínio. Dessa forma, seguindo estes critérios, as fontes secundárias que foram produzidas fora de um contexto histórico em especial, distanciam-se do tipo de documento que se pretende representar na ontologia.

5.3.8 Algumas considerações sobre as classes e as propriedades

As análises feitas sobre cada classe acima teve por finalidade refletir sobre questões que versam sobre a representação dos conjuntos de termos encontrados no domínio da Historia. Subclasses como “Papel Processual” e “Papel Social”, ou ainda “Regiao Politica Administrativa”, “Espaço Construído” e “Espaço Natural” foram sugeridas para fins de organização mental prévia das classes adotadas, justificando sua entrada e abrangência dos conceitos. Dessa forma, quando instanciarmos a ontologia com assertivas que apontam que Carlos Lacerda (pessoa) foi vítima (papel processual) de atentado (evento) em Copacabana (bairro), teremos simplesmente a associação de uma instância da classe Agente a instâncias da classe Papel, Evento e Lugar.

O quadro-síntese a seguir apresenta algumas propriedades possíveis para as instâncias das classes discutidas, com uma proposta de definição de seu emprego, domínio e escopo, além de referências semelhantes em outras ontologias. É basicamente uma seleção dos relacionamentos listados acima; uma pequena amostra de como deve ser conduzida a etapa de conceitualização de uma ontologia, segundo as orientações propostas nas metadologias visitadas.

Para um levantamento mais completo destes relacionamentos, é imprescindível debruçarmo-nos com mais profundidade nas teorias para o estabelecimento das relações conceituais advindas da Terminologia, da Ciência da Informação e Ciência da Computação, que visitamos na seção 2.2.3 do capítulo que versa justamente sobre os sistemas de conceitos e as teorias classificatórias no campo da Representação do Conhecimento.

Tabela 9 - Quadro-síntese de definições dos relacionamentos:

Relacionamento	Definição					
	Tipos de Propriedades	Cardina-lidade	Domínio *(Domain)*	Escopo *(Range)*	Propriedade inversa	Referências em outras ontologias
temLocalizacaoEspacial	Associa um evento, agente ou mesmo um lugar a uma localização no espaço. Ex: 'o presidente estava no Palácio do Catete', 'o monumento aos pracinhas localiza-se no Rio de Janeiro'. A entidade só pode estar em um lugar por vez, mas um lugar pode ser a localização espacial de várias entidades. Por isso a cardinalidade 1 para N.					
	Funcional	1, N	Evento, Agente, Objeto, Lugar	Lugar	ehLocalizacaoEspacialDe	dul:hasLocation edm:happenedAt cidoc:tookPlaceAt
temLocalizacaoTemporal	Associa um evento ao tempo de ocorrência. Um evento só pode ocorrer em um dado tempo.					
	Funcional	1, N	Evento	Tempo	ehLocalizacaoTemporalDe	edm:occurredAt
temParticipante	Associa pessoas, coisas ou recursos a um evento no qual eles estiveram presente. Ex: 'no Atentado da Tonelero, Lacerda participou como vítima', 'o assassinato foi cometido/contou com a participação de uma arma de fogo'.					
	anti-simétrica	N, N	Evento	Agente, Objeto	participouDe	dul:hasParticipant edm:wasPresentAt cidoc:hadParticipant
temPapel	Relação entre um agente e o papel que desempenha. Um agente pode desempenhar ou ter mais de um papel.					
	anti-simétrica	N, N	Agente, Objeto	Papel	ehPapelDe	dul:isRoleOf
temMembro	Relação entre entidades da classe Agente. Ex: 'uma pessoa X é membro de um grupo Y', 'organização Z é membro de outra organização ZZ'.					
		N, N	Agente	Agente	ehMembroDe	dul:hasMember
ehConstituidoDe	Relação partitiva entre instâncias das classes do domínio. Ex: 'século 20 contém o ano 1954', 'II Guerra Mundial contém o evento da conquista de Monte Castelo'.					
	anti-simétrica transitiva	1, N	Tempo, Lugar, Evento, Objeto	Tempo, Lugar, Evento, Objeto	estahContidoEm	dul:hasPart cidoc:isComposedOf

trabalhaPara	Relação entre agentes, onde é estabelecida uma situação profissional ou de delegação					
		N, N	Agente	Agente	trabalhaAtravesDe	dul:actsFor
ehSobre	Relação entre um documento e uma outra entidade. É reflexiva porque documentos também podem ser sobre eles mesmos.					
	Reflexiva	N, N	Documento	Pessoa, Lugar, Objeto, Documento, Tempo	temReferencia	dul:isAbout cidoc:appliesTo
ehAutorDe	Relacao entre um Agente e um Objeto, usualmente um Documento					
	anti-simétrica	N, N	Agente	Documento	temAutor	
ehReferenciaDe	Relação entre um Documento e uma entidade, onde o primeiro constitui informação descritiva sobre o segundo.					
		N, N	Documento	Agente, Evento, Objeto, Papel, Lugar PeriodoHistorico	ehReferenciadoPor	dul:describes
conceitua	Relação que diz que alguém (um agente) toma para si uma dada visão sobre um conceito, evento, etc. Ex: 'Vargas *acredita* na teoria da conspiração', 'para a imprensa, o assassinato *parece* ter sido planejado por algum político getulista'.					
		1, N	Agente	Agente, Evento, Objeto, Documento, PeriodoHistorico	ehConceituadaPor	dul:conceptualizes
sucede	Relação de ordenamento entre um evento ou uma entidade de tempo. Ex: 'Era Vargas sucede Primeira República', 'instauração do regime militar sucede golpe de 1964', 'ano 2000 sucede 1999'.					
	transitiva anti-simétrica		Evento, Data, PeriodoHistorico	Evento, Data, PeriodoHistorico	precede	dul:follows
encontrou	Associa pessoas a outras pessoas ou coisas que elas encontraram					
		N, N	Agente	Agente, Objeto	foiEncontrado	edm:hasMet

5.3.9 Exemplo de instanciação das classes propostas

O mapa conceitual apresenta um exemplo de instanciação simples das classes propostas na ontologia HIST.

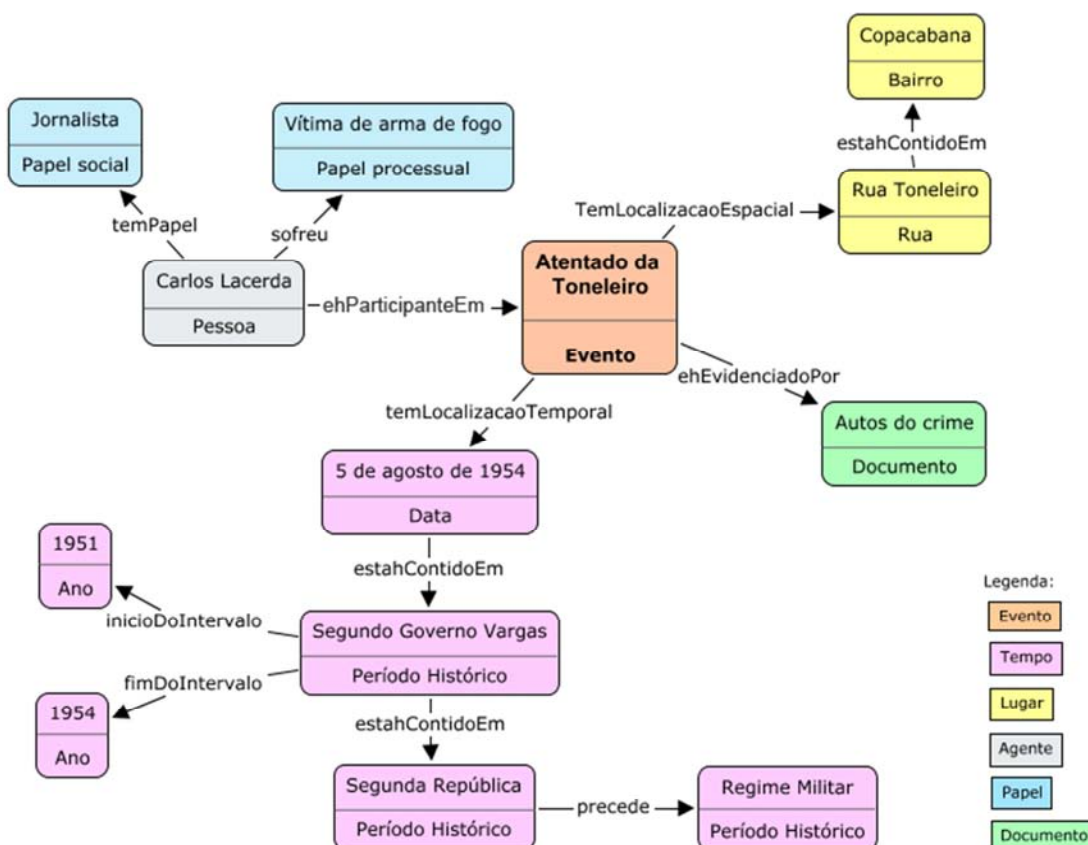



Figura 54 - Exemplo de instanciação das classes propostas em HIST

Inúmeras outras propriedades podem ser arroladas para dar sustentação à semântica do domínio. Como já mencionado, Guarino e Welty (2004), assim como Guizzardi (2005) e Guizzardi & Wagner (2005, 2010) propõem métodos e modelos para validação da adequação dos relacionamentos taxonômicos em ontologias, baseados principalmente em noções filosóficas de identidade, unidade e individualidade. Estas metodologias foram apresentadas ao longo da dissertação e sua importância é indiscutível, devendo ser consideradas em um plano de aplicação real.

Com as classes e propriedades definidas, o passo de formalização pode ser feito com a ajuda de ferramentas próprias para isso. A maioria dos editores de ontologia traz uma interface intuitiva para a estruturação das classes, a criação de regras, de relacionamentos, e

inclusão de instâncias. Permitem também realizar algumas inferências e consultas por sobre as classes e entidades, além de aceitar a exportação para as linguagens de representação XML, RDF, OWL e outros. A figura abaixo apresenta a interface do Protegé[91], um editor de ontologias bastante utilizado e conhecido pela simplicidade de uso, e uma visão parcial das classes de nossa ontologia.

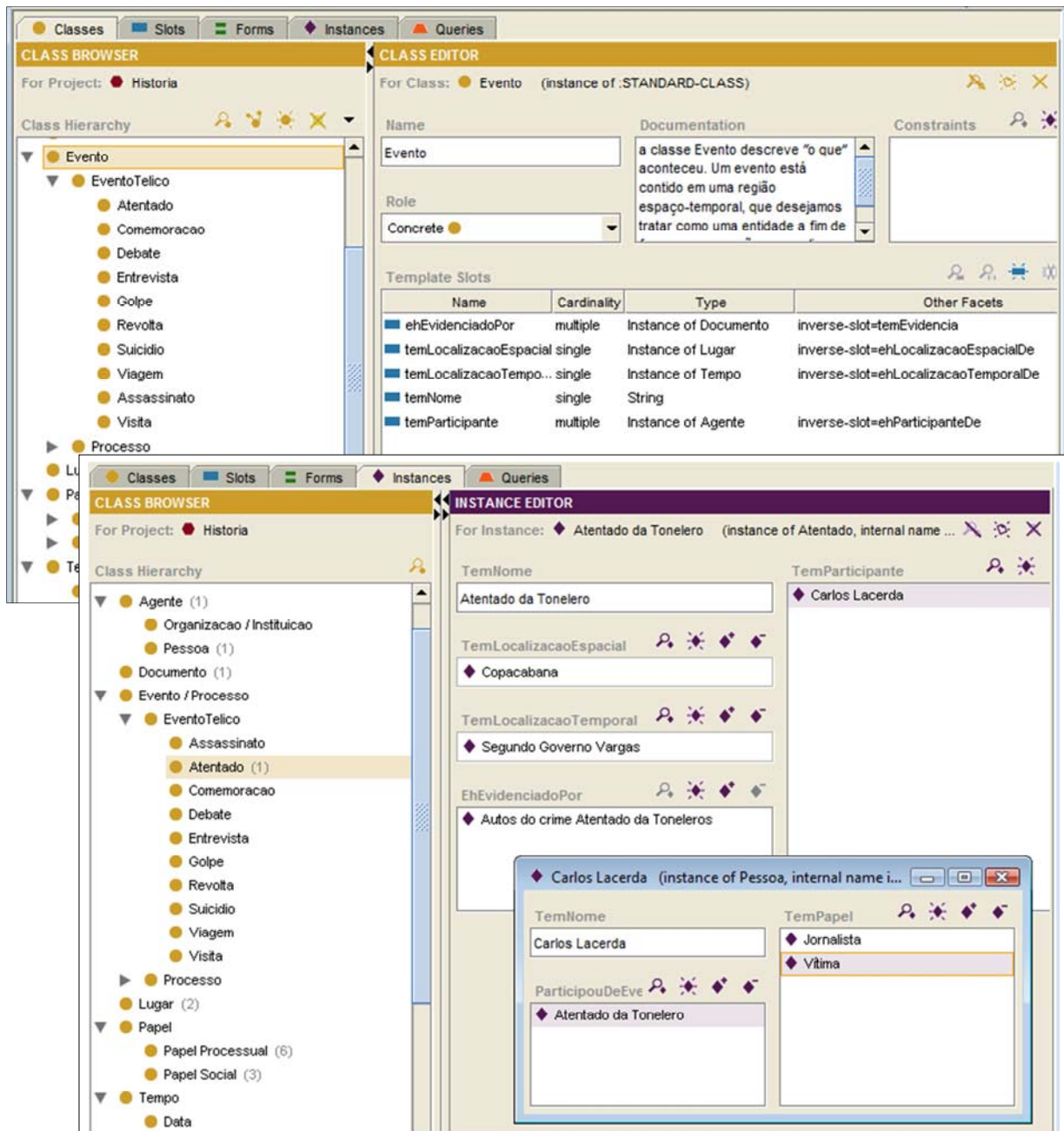



Figura 49 - Interfaces do editor de ontologias Protegé

---

[91] http://protege.stanford.edu/

6 Considerações finais e possibilidades futuras

À luz do exposto sobre a Representação do Conhecimento, observa-se que este não é um campo acadêmico novo. Renova-se pela disponibilidade de estudos e inovações tecnológicas, mas o arcabouço teórico-conceitual vem se construindo há séculos.

Este estudo buscou investigar as bases teóricas e metodológicas existentes nesse campo da representação do conhecimento para a modelagem de uma ontologia no domínio da História, identificando seus conceitos nucleares. Nesse sentido, consideramos o objetivo alcançado principalmente através dos capítulos 2, *Fundamentos teóricos* e 5, *Proposta de modelagem conceitual da ontologia*. O primeiro, ao abordar os aspectos que fundamentam questões importantes sobre a representação do conhecimento, incluindo as noções de modelos conceituais e ontologia. E o último, por derivar as categorias que consideramos primordiais para a representação de conceitos na História.

A fase de conceitualização é em geral seguida pela fase de formalização do modelo, porém esta não é contemplada no estudo, pois está além do objetivo proposto. Apesar disso, seus requisitos foram levantados durante a revisão das metodologias para, certamente, constar em trabalhos futuros. Alguns exemplos de restrições sobre o domínio expressas em OWL puderam ser conferidas no capítulo que versa sobre as linguagens de representação, em 2.3.3, e é uma amostra do que se espera, por extensão, na continuidade da pesquisa.

Não é tarefa fácil chegar a um consenso sobre a melhor definição para um conceito ou que características de divisão usar para formar as cadeias e renques. Este é, aliás, um processo que suscita muitas controvérsias a nível epistemológico sobre o domínio, conforme demonstram as reuniões realizadas pelo grupo de trabalho sobre Tesauro de Regime Militar Brasileiro[92] da FGV. Nelas, cada conceito é analisado pelos seus membros, em densas discussões. E muitas vezes, decidem-se por critérios que evitam a inclusão de termos delicados ou que poderiam ser ética ou politicamente mal interpretados. São decisões tomadas em consenso.

Descobrimos por experiência o que é fato na literatura revisada: de que a modelagem conceitual de uma dada porção da realidade carrega o sentido de verdade somente para os

---

[92] Grupo de trabalho conduzido no interior do Centro de Pesquisa e Documentação de História Contemporânea do Brasil em parceria com a Escola de Matemática Aplicada, ambos da Fundação Getulio Vargas.

compromissos adotados para aquela realidade. Ou seja, porque um conceito pode ser definido de *n* maneiras, é necessário explicitar claramente as escolhas tomadas e seus contextos de uso para que o modelo faça sentido. Diferentemente do que parece ocorrer nas ciências ditas naturais, onde é mais fácil dizer, por exemplo, que uma célula é uma célula ou uma função molecular tem esta ou aquela característica, na história um personagem pode ser identificado como herói, mártir e traidor ao mesmo tempo, sendo todas estas concepções verdadeiras nos seus distintos contextos de uso.

Nesse sentido a fase de conceitualização parece ser uma das mais difíceis na construção de ontologias de história porque envolve reconhecer como as pessoas percebem o mundo e como elas categorizam coisas em suas mentes. Carr (1978, p. 56) aponta que “a história é necessariamente subjetiva, pois é o homem que se observa a si próprio”. Para Talja (1997), “conhecimento e estruturas de conhecimento não são nem objetivas nem subjetivas, mas intersubjetivas, produzidas no interior de um sistema de significados compartilhados” (TALJA, 1997).

Além disso, quando formulamos conceitos, é importante evitar as armadilhas de se elevar as características individuais como se elas fossem típicas ou gerais, ou generalizar excessivamente criando categorias a partir de muito poucos exemplares.

A História, como todas as áreas de humanidades, possui aspectos desafiadores porque não se fia em classificações terminológicas rigorosas como acontece com a biologia, genética e medicina. Tradicionalmente, estes domínios trazem informações muito mais estruturadas (em termos de bancos de dados, produção científica, etc.) demandadas pela própria dinâmica de pesquisa, cujas descobertas necessitam ser compartilhadas de forma rápida e eficiente. Ao contrário do que acontece na História.

Como trabalho futuro, podemos citar uma revisão mais detalhada de outros sistemas de classificação, como as Classificações Decimais de Dewey e Universal, a Colon Classification de Ranganathan, e o Library Congress Classification da Biblioteca do Congresso americano, que, apesar de bibliográficos, podem ajudar nas escolhas das facetas, sendo também boa fonte teórica relativa à análise e síntese dos assuntos. A definição de classes completas com a criação de um conjunto de condições necessárias e suficientes é primordial para o desenvolvimento da ontologia para aplicação no mundo real. Dessa forma, será possível

computar a hierarquia de classes através de um mecanismo de inferência, onde subclasses são inferidas automaticamente desde que satisfaçam as condições estabelecidas.

Os resultados do estudo encorajam a continuação do processo de criação, dessa vez com a ajuda sistemática dos especialistas da área. São eles que terão a intuição, prerrogativas e conhecimentos necessários para a validação das escolhas do modelo rumo a uma ontologia capaz de representar adequadamente nosso passado recente. A possibilidade de se realizar colheitas de taxonomias para tipologias da História ajudaria a preencher uma lacuna que existe hoje no domínio, e serviria como ponto de partida para futuras expansões.

Conforme as pesquisas no campo das humanidades progridam, a nossa ontologia HIST poderia se combinar a iniciativas semelhantes, dando forma a um esquema modelo, tal como o fornecido pelo CIDOC CRM ou o Europeana.

Enfim, as dificuldades existem e os desafios são muitos, mas as possibilidades abertas por este cenário de inovação levam a um contínuo e merecido esforço para tentar superá-los. A interdisciplinaridade é traço inerente ao conjunto de iniciativas, e certamente a CI, ciência que, por natureza, possui esta característica e coloca como central a questão da informação desde a sua gênese até o processo de transformação de dados em conhecimento, tem papel fundamental neste cenário.

“Se queremos progredir, não devemos repetir a história,
mas fazer uma história nova”

(Mahatma Gandhi)

7 Referências

AKKER, C.; AROYO, L, CYBULSKA, A; ERP, M; GORGELS, P; HOLLINK, L; JAGER, C; LEGENE, S; MEIJ, L; OOMEN, J; OSSENBRUGGEN, J; SCHREIBER, G; SEGERS, R; VOSSEN, P; WIELING, B. *Historical Event-based Access to Museum Collections* In: Proceedings of the First International Workshop on Recognising and Tracking Events on the Web and in Real Life (EVENTS2010), Athens, 2010. Disponível em: http://ceur-ws.org/Vol-624/paper1.pdf

ALLEMANG, Dean & HENDLER, James. Semantic Web for the Working Ontologist: effective modeling in RDFs and OWL. Morgan Kauffman Publ, 2008.

ALLEN, James; FERGUSON, George. Actions and events in interval temporal logic. Technical Report. University of Rochester, New York, 1994. Disponível em: http://web.mit.edu/larsb/Public/16.412/pset%204/allen94actions.pdf

ALMEIDA, Maurício B. Um modelo baseado em ontologias para representação da memória organizacional. Tese. UFMG, Belo Horizonte, 2006. Disponível em: http://hdl.handle.net/1843/VALA-6T7QFT. Acesso em: 27/06/2010.

ALMEIDA, Maurício B; BAX, Marcelo. *Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção*. Revista Ciência da Informação, 32(3), 2003. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652003000300002&lng=en&nrm=iso. Acesso em 07/08/2010.

ALMEIDA, Maurício B; SOUZA, Renato Rocha. *Avaliação do espectro semântico de instrumentos para organização da informação* Encontros Bibli, v. 16, p. 25-50, 2011. Disponível em: http://www.periodicos.ufsc.br/index.php/eb/article/view/11963. Acesso em 23/04/2012.

ALVARENGA, Lídia. *Representação do conhecimento na perspectiva da Ciência da Informação em tempo e espaços digitais*. Encontros Bibli, n. 15, Universidade Federal de

Santa Catarina, 2003. Disponível em: http://redalyc.uaemex.mx/redalyc/pdf/147/14701503.pdf. Acesso em 05/12/2010.

AZEVEDO NETO, Carlos Xavier. *A abordagem do conceito como uma estrutura semiótica*. Transformação, Campinas, n. 20, jan./abr. 2008. Disponível em: http://www.brapci.ufpr.br/download.php?dd0=13023. Acesso em 03/01/2011.

BARROS, José. O Campo da História. Petrópolis: Vozes, 2009.

BENJAMINS, V. Richard; RADOFF, Mark; DAVIS, Mike; GREAVES, Mark; LOCKWOOD, Rose; CONTRERAS, Jesús. *Semantic technology adoption: a business perspective*. In: DOMINGUE, John; FENSEL, Dieter; HENDLER, James (eds). Handbook of Semantic Web Technologies. Springer-Verlag Berlin Heidelberg, 2011.

BENNETT, Brandon; GALTON, Antony. *A Unifying Semantics for Time and Events.* Artificial Intelligence, volume 153, no. 1-2. 2004. Disponível em http://www.sciencedirect.com/science/article/pii/S0004370203001681. Acesso em 05/05/2011.

BERNERS-LEE, Tim. Semantic Web - XML2000. 2000. Disponível em: http://www.w3.org/2000/Talks/1206-xml2k-tbl. Acesso em: 18/10/2010.

BERNERS-LEE, Tim. Semantic Web road map. 1998. Disponível em: http://www.w3.org/DesignIssues/Semantic.html . Acesso em 18/10/2010.

BERNERS-LEE, Tim; HALL, Wendy; SHADBOLT, Nigel. *The Semantic Web Revisited*. IEEE Intelligent Systems, 21, n 3, May/Jun 2006, pp. 96-101. Disponível em: http://eprints.ecs.soton.ac.uk/12614. Acesso em 31/10/2010.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. *The Semantic Web*. Scientific American, May 2001.

BioPortal. VIVO Ontology. [Versão de] 20 de dezembro de 2011. Disponível em: http://purl.bioontology.org/ontology/vivo

BIROU, Alain. Dicionário das Ciências Sociais. (5ª ed.). Lisboa: Dom Quixote, 1982.

BLAIR, David C. Wittgenstein, Language and Information: "Back to the Rough Ground". Springer, 2006.

BLOCH, Marc. Apologia da história, ou, O ofício do historiador. Rio de Janeiro: Jorge Zahar Editor, 2001.

BONTCHEVA, Kalina; CUNNINGHAM, Hamish. *Semantic annotatios and retrieval: manual, semiautomatic and automatic generation*. In: DOMINGUE, John; FENSEL, Dieter; HENDLER, James A (eds). Handbook of Semantic Web Technologies. Springer-Verlag Berlin Heidelberg, 2011.

BOOCH, Grady; JACOBSON, Ivar; RUMBAUGH, James. UML Guia do Usuário. Editora CAMPUS. Rio de Janeiro. 2006.

BORST, W.N. Construction of Engineering Ontologies for Knowledge Sharing and Reuse. Tese (Phd). University of Tweenty, 1997. Disponível em: http://www.ub.utwente.nl/webdocs/inf/1/t0000004.pdf. Acesso em: 20/12/2010.

BORKO, H. Information Science: *What is it?* American Documentation, v. 19, n. 1, p.3-5. 1968

BOX-SEFFENSMEIER, Janet M.; JONES, Bradford S. Event History Modeling: a guide for social scientists (Analytical Methods for Social Research), 2004.

BRACHMAN, Ronald J; LEVESQUE, Hector J. Knowledge representation and reasoning. Elsevier, 2004.

BRAGA, Gilda Maria. *Informação, ciência da informação: breves reflexões em três tempos*. Ciência da Informação – V. 24, número 1, 1995.

BRASCHER, Marisa; CAFÉ, Lígia. *Organização da Informação ou Organização do Conhecimento?* Comunicação oral apresentada no IX Enancib. USP, São Paulo, 2008. Disponível em: http://www.dgz.org.br/jun09/Art_03.htm. Acesso em 03/01/2011.

BRESLIN, John; DAVIS, Mills; NOVA, Spivak. *What is the Evolution of the Internet to 2020?* In Project10X's Semantic Wave 2008 Report: Industry Roadmap to Web 3.0 & Multibillion Dollar Market Opportunities. Disponível em: http://www.isoco.com/pdf/Semantic_Wave_2008-Executive_summary.pdf. Acesso em 20/10/2010.

BURKE, Peter. *A nova história, seu passado e seu futuro*. In: BURKE, Peter (org). A escrita da história: novas perspectivas. São Paulo: Ed. Unesp. 1992.

BUSH, Vannevar. *As we may may think*. Atlantic Monthly, v.176, n.1, p. 101-108, 1945. Disponível em: http://www.cs.sfu.ca/CC/365/mark/material/notes/Chap1/VBushArticle/ . Acesso em 25/05/2010.

CAMPOS, Linair Maria. Diretrizes para definição de recorte de domínio no reuso de ontologias biomédicas: uma abordagem interdisciplinar baseada na análise do compromisso ontológico. Tese. Ciência da Informação. Universidade Federal Fluminense, Niterói, RJ. 2011.

CAMPOS, Maria Luiza de A. Linguagem Documentária: Teorias que fundamentam sua elaboração. Niterói, RJ: EUFF, 2001.

CAMPOS, Maria Luiza de *A. Modelização de Domínios de Conhecimento: uma investigação de princípios fundamentais*. Ciência da Informação, Brasília, v. 33, n. 1, 2004. Disponível em: <http://www.ibict.br/cionline/viewarticle.php?id=77&layout=abstract>. Acesso em 03/05/2010.

CAMPOS, Maria Luiza de A.; BRASIL, Maria Irene; COELHO, Beatriz A. S.; BASTOS, Dilza R. *Vocabulário Sistematizado: a experiência da Fundação Casa de Rui Barbosa*. In: INTEGRAR, 2002, São Paulo. INTEGRAR. São Paulo, 2002.

CAMPOS, Maria Luiza de A.; GOMES, Hagar Espanha. *Taxonomia e Classificação: princípios de categorização*. Datagramazero (Rio de Janeiro), v. 9, n. 1, 2008. Disponível em: http://www.dgz.org.br/ago08/Art_01.htm Acesso em 10/05/2010.

CAPURRO, Rafael. *Epistemologia e Ciência da Informação*. In: ENANCIB – Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação, 5, Belo Horizonte: ECI/UFMG, 2003.

CAPURRO, Rafael. *What is information science for? A philosophical reflection*. In: Vakkari, Perti, Cronin, Blaise. Conceptions of library and information science. Tempere, Taylor Graham, 1991.

CARDOSO, Ciro Flamarion. *História e paradigmas rivais*. In: CARDOSO, Ciro Flamarion; VAINFAS, Ronaldo. Domínios da história: ensaios de teoria e metodologia. Rio de Janeiro: Campus, 1997.

CARR, Edward Hallet. *Que é história?*. Conferências George Macaulay Trevelyan, Universidade de Cambridge. Paz e Terra, 1978.

CASATI, Roberto; VARZI, Achille C. *Event concepts*. In: Understanding Events: How Humans See, Represent, and Act on Events. New York: Oxford University Press, 2008. Disponível em <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.72.701&rep=rep1&type=pdf>. Acesso em 06/06/2010.

CHIERCHIA, Gennaro. Semântica. (trad. Luiz Arthur Pagani; Ligia Negri; Rodolfo Ilari). Campinas, SP: Editora da Unicamp, 2003.

COHEN, Daniel J. "*History and the Second Decade of the Web*". Rethinking History. 8.2 293-301, 2004. Disponível em: http://www.dancohen.org/files/hist_2nd_decade_web.pdf

COHEN, Daniel J. "*Social and Semantic Computing for Historical Scholarship*". In History and Technology. Perspectives, 2007. Disponível em: http://www.historians.org/Perspectives/issues/2007/0705/0705tec2.cfm

CORCHO, Oscar; FERNÁNDEZ-LÓPES, Mariano; GÓMEZ-PÉREZ, Assunción. *Methodologies, tools and languages for building ontologies: where is the meeting point*? In: Data & Knowledge Engineering 46. Elsevier, 2003. Disponível em:

http://www.dia.fi.upm.es/~ocorcho/documents/DKE2003_CorchoEtAl.pdf. Acesso em 20/12/2010.

CORNELIUS, I. *Information and Interpretation*. In: Peter Ingwersen, Niels O. Pors (Eds.) Proceedings CoLIS2. Second International Conference on Conceptions of Library and Information Science: Integration in Perspective. October 13-16, 1996. The Royal School of Librarianship, Copenhagen, 11-21.

COUGO, Paulo Sérgio. Modelagem conceitual e projeto de banco de dados. Rio de Janeiro: Campus. 1997.

(CPDOC) ABREU, Alzira A; BELOCH, Israel; LAMARAO, Sergio. T. N.; LATTMAN-WELTMAN, Fernando; PAULA, Christiane. J. (orgs). Dicionário Histórico-Biográfico Brasileiro Pós-1930 , 2010.

CYBULSKA, Agata; VOSSEN, Piek. *Event models for historical perspectives: determining relations between high and low level events in text, based on the classification of time, location and participants*. Lrec2010, Proceedings, 2011. Disponível em: http://www.lrec-conf.org/proceedings/lrec2010/pdf/205_Paper.pdf. Acesso em 20/06/2012.

D'ARAUJO, Maria Celina. A Era Vargas. 1a ed. São Paulo: Moderna, 1997.

DAHLBERG, I. Ontical structures and universal classification. Bangalore: Sarada Ranganthan Endowment,1978.

DAHLBERG, I. *Teoria da classificação, ontem e hoje*. In: CONFERÊNCIA BRASILEIRA DE CLASSIFICAÇÃO BIBLIOGRÁFICA. Rio de Janeiro: Conferência Brasileira de Classificação Bibliográfica, 1976.

DAHLBERG, Ingetraut. *Conceptual structures and Systematization*. International Forum on Information and Documentation, v. 20, n. 3, 1995. p. 9-24.

DAVIDSON, D. Essays on Action and Events. Oxford, 1980.

DAVIS, R; SHROBE, H; SZOLOVITS, P. *What is a Knowledge Representation*? AI Magazine , 14(1):17-33, 1993.

DESWARTE, R. e OOSTHOEK, J. *Creating a history ontology: The challenges and criteria of building a humanities ontology*. In Conference National Centre for e-Social Science 2005 (NCeSS), 2005.

DODEBEI, Vera Lucia Doyle. Tesauro: linguagem de representação da memória documentária. Niterói; Rio de Janeiro: Intertexto; Interciência, 2002.

DOMINGUE, John; FENSEL, Dieter; HENDLER, James A.*Introduction to the Semantic Web Technologies*. In: DOMINGUE, John; FENSEL, Dieter; HENDLER, James A (eds). Handbook of Semantic Web Technologies. Springer-Verlag Berlin Heidelberg, 2011.

DUFF, W.M.; JOHNSON, C.A. *Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives*. The Library Quarterly, 72(4). 2004. Disponivel em: http://www.jstor.org/stable/10.2307/40039793

FELBER, H. *The Vienna School of Terminology: fundamentals and its theory*. In: INTERNATIONAL SYMPOSIUM ON THEORETICAL AND METHODOLOGICAL PROBLEMS OF TERMINOLOGY, 1979, Moscow. Proceedings... Muenchen : Saur, 1981.

FERNANDEZ, Rosali P. *Classificação: Um processo fundamental da natureza humana*. In: CONFERÊNCIA BRASILEIRA DE CLASSIFICAÇÃO BIBLIOGRÁFICA, 1976, Rio de Janeiro. Anais... Rio de Janeiro: Conferência Brasileira de Classificação Bibliográfica, 1976.

FERNÁNDEZ-LÓPEZ, M., GOMEZ-PÉREZ, A. *Searching for a Time Ontology for Semantic Web Applications*. In: Formal Ontology in Information Systems, Turín, Italy (2004). Disponível em: http://oa.upm.es/5493/1/Searching_for_a_Time_Ontology_.pdf. Acesso em 21/02/2011.

FERREIRA, Marieta de Moraes. *História do tempo presente: desafios*. Cultura Vozes, Petrópolis, v.94, nº 3, p.111-124, maio/jun., 2000.

FERREIRA, Marieta de Moraes. Apresentação. In: CPDOC 30 anos. Rio de Janeiro, Editora FGV. 2003.

FODOR, Jerry. *Semântica: uma entrevista com Jerry Fodor*. ReVEL. Vol. 5, n. 8, 2007. Disponível em http://www.revel.inf.br/site2007/_pdf/8/entrevistas/revel_8_entrevista_jerry_fodor.pdf. Acesso em 05/11/2010.

FOUCAULT, Michel. A arqueologia do saber. Trad. L. F. Baeta Neves. Petrópolis: Vozes. 1972.

FOWLER, Martin. UML distilled: a brief guide to the standard object modeling language. Boston, Addison Wesley, 2003.

FREITAS, Lídia Silva de. *O dispositivo de arquivo: a construção histórico-discursiva do documento e do fato*. In: FREITAS, Lidia Silva de; MARCONDES, Carlos H.; RODRIGUES, Ana Célia (Orgs.). DOCUMENTO: gênese e contextos de uso. Niterói: PPGCI, EDUFF, 2010. p. 141-160.

FROHMANN, B. *Rules of indexing: a critique of mentalism in information retrieval theory*. Journal of documentation. London: Aslib, 1990, v. 46, n. 2, p. 81-101.

GANGEMI, Aldo; GUARINO, Nicola; MASOLO, Claudio; OLTRAMARI, Alessandro; SCHNEIDER, Luc. *Sweetening Ontologies with DOLCE*. Springer, 2002. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.11.6038

GIL, Fernando. *Categorizar*. In: ENCICLOPÈDIA Einaudi. Conhecimento. [s.l.]: Imprensa Nacional – Casa da Moeda, p.52-89, v. 41, 2000.

GIUNCHIGLIA, Fausto; ZAIHRAYEU, Ilia. *Lightweight Ontologies*. Technical Report DIT-07-071. University of Trento, Italia, 2007. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.1717&rep=rep1&type=pdf. Acesso em 18 de maio de 2012.

GOMES, Hagar Espanha (Org.). Manual de elaboração de tesauros monolíngues. Brasília: Programa Nacional de Bibliotecas de Instituições de Ensino Superior, 1990.

GOMEZ-PEREZ, A. *Ontological Engineering: A state of the art*. Expert Update. British Computer Society. v. 2, n. 3, p. 33-43. 1999. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.28.8567&rep=rep1&type=pdf Acesso em 20/12/2010.

GOMEZ-PEREZ, A.; FERNANDEZ, M.; JURISTO, H. Methontology: from ontological art towards ontological engineering. 1997. Disponível em: <http://www.aaai.org/Papers/Symposia/Spring/1997/SS-97-06/SS97-06-005.pdf>. Acesso em: 20/12/2010.

GOMEZ-PEREZ, A; FERNANDEZ, M; VICENTE, A J. Towards a method to conceptualize a domain ontology. 1996. Disponível em http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.24.167&rep=rep1&type=pdf Acesso em 20/12/2010.

GONZÁLEZ, José Moreiro. *Palavra, termo, conceito: das linguagens documentárias até os vocabulários semânticos para a web*. In: FREITAS, Lidia Silva de; MARCONDES, Carlos H.; RODRIGUES, Ana Célia (Orgs.). DOCUMENTO: gênese e contextos de uso. Niterói: PPGCI, EDUFF, 2010.

GRACIOSO, Luciana de Souza & SALDANHA, Gustavo. Ciência da Informação e Filosofia da Linguagem: da pragmática informacional à web pragmática. 2010. pp. 99 a 118.

GRAHAM, S. R. *Historians and Electronic Resources: Patterns and Use*. Journal of the Association for History and Computing, 5(2). 2002. Disponível em: http://mcel.pacificu.edu/JAHC/JAHCV2/ARTICLES/graham/graham.html

GREENBERG, Jane; SUTTON, Stuart; CAMPBELL, D Grant. *Metadata: a Fundamental Component of the Semantic Web*. Bulletin of The American Society for Information Science and Technology, v. 29, n. 4, Apr/May 2003, pp 16-18.

GRIMM, Stephan; ABECKER, Andreas; VOLKER, Johanna; STUDER, Rudi. *Ontologies and the Semantic Web*. In: DOMINGUE, John; FENSEL, Dieter; HENDLER, James A (eds). Handbook of Semantic Web Technologies. Springer-Verlag Berlin Heidelberg, 2011.

GRUBER, Tom. *Collective Knowledge Systems: Where the Social Web meets the Semantic Web*. To appear in Journal of Web Semantics, 2007. Disponível em http://tomgruber.org/writing/collective-knowledge-systems.htm. Acesso em 23/10/2010.

GRUBER, Tom. Ontologies. 2008. Disponível em http://tomgruber.org/writing/ontology-definition-2007.htm. Acesso em 23/10/2010.

GRUBER, Tom. *Towards principles for the design of ontologies used for knowledge sharing*. . Int. J. Human-Computer Studies, v. 43, n. 5/6, 1993.

GRUNINGER, Michael; FOX, Mark S. *Methodology for the Design and Evaluation of Ontologies*: Proceedings of the Workshop on basic Ontological Issues in Knowledge Sharing. IJCAI-95, Canadá, 1995.

GUARINELLO, Norberto Luiz. *História científica, história contemporânea e história cotidiana*. Rev. Bras. Hist., São Paulo, v. 24, n. 48, 2004 .

GUARINO, N. *Some ontological principles for designing upper-level lexical resources.* Proceedings of the First International Conference on Language Resources and Evaluation (pp. 527-534). Granada, Spain: European Language Resources Association, 1998. Disponível em: ww.loa.istc.cnr.it/Papers/LREC98.pdf

GUARINO, N.; WELTY, C. *Evaluating ontological decisions with OntoClean*. Communications of the ACM, v. 45 , n. 2, Feb. 2002. Disponível em: http://portal.acm.org/citation.cfm?id=503150.

GUARINO, Nicola, WELTY, Chris, *An Overview of OntoClean*, in STAAB, S. STUDER, R. (eds.), Handbook on Ontologies, Springer Verlag, 2004. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.7618&rep=rep1&type=pdf. Acesso em 18/05/2010.

GUARINO, Nicola. *Formal Ontology and Information Systems*. Proceedings of FOIS'98, Trento, Italy, 1998. Disponível em: http://fois08.dfki.de/proceedings/FOIS08PosterProceedings.pdf. Acesso em 18/05/2010.

GUARINO, Nicola. *Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration*. In M. T. Pazienza (ed.) Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology. Springer Verlag: 139-170. 1997a. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.29.1776&rep=rep1&type=pdf. Acesso em 18/05/2010.

GUARINO, Nicola. *Some Organising Principles For A Unified Top Level Ontology*. In Working Notes of AAAI Spring Symposium on Ontological Engineering, Stanford, USA. 1997. Disponível em: http://www.aaai.org/Papers/Symposia/Spring/1997/SS-97-06/SS97-06-008.pdf. Acesso em 18/05/2010.

GUARINO, Nicola; GIARETTA, Pierdaniele. *Ontologies and Knowledge Bases: Towards a Terminological Clarification*. In: N.J.I. MARS (Ed.) Towards very large Knowledge Bases. IOS Press, Amsterdam, 1995.

GUIZZARDI, G., FALBO, R. A., GUIZZARDI, R. S. S. *A importância de Ontologias de Fundamentação para a Engenharia de Ontologias de Domínio: o caso do domínio de Processos de Software*. IEEE Transactions Latin America, 2008.

GUIZZARDI, Giancarlo. *An Introduction to Ontology Engineering*. Tutorial. I Conferência Web W3C Brasil. São Paulo, 2009. Disponível em http://conferenciaweb.w3c.br/2009/trabalhos/w3c2009-sp_gg.ppsx . Acesso em 20/10/2010.

GUIZZARDI, Giancarlo. Desenvolvimento para e com reuso: um estudo de caso no domínio de vídeo sob demanda. Dissertação. Universidade Federal do Espirito Santo, 2000. Disponível em: http://www.loa-cnr.it/Guizzardi/MSc.htm. Acesso em 25 de outubro de 2010.

GUIZZARDI, Giancarlo. Ontological Foundations for Structural Conceptual Models, PhD Thesis (CUM LAUDE), University of Twente, The Netherlands. Published as the book

“Ontological Foundations for Structural Conceptual Models”, Telematica Instituut Fundamental Research Series No. 15. 2005. Disponível em: http://www.inf.ufes.br/~gguizzardi/OFSCM.pdf. Acesso em 08 de setembro de 2010.

GUIZZARDI, Giancarlo; WAGNER, Gerd. *Towards an ontological foundation of discrete event simulation*. Proceedings of the 2010 Winter Simulation Conference. 2010. Disponível em: http://www.inf.ufes.br/~gguizzardi/059.pdf

GUIZZARDI, Giancarlo; WAGNER, Gerd. *Some Applications of a Unified Foundational Ontology in Business Modeling*. Ontologies and Business Systems Analysis, Michael Rosemann and Peter Green (Eds.), IDEA Publisher, 2005. Disponível em: http://www.loa-cnr.it/Guizzardi/book.pdf. Acesso em 08 de setembro de 2010.

GUIZZARDI, Giancarlo; WAGNER, Gerd. *Using the Unified Foundational Ontology (UFO) as a Foundation for General Conceptual Modeling Languages*. In: Theory and Application of Ontologies ed.Berlim: Springer-Verlag, 2010.

VAN HEIJST, G.; SCHREIBER, A. Th; WIELINGA, B.J. *Using explicit ontologies in KBS development*, Intl. J. Human-Computer Studies 46, 1997. Disponível em: http://www.cs.vu.nl/~guus/papers/Heijst97a.pdf. Acesso em 30/05/2012.

HENDLER, James. *The Semantic Web: Agents and the Semantic Web*. IEEE Intelligent Systems, 2001. Disponível em: http://hcs.science.uva.nl/Capita-AI/2002/papers/hendler.pdf. Acesso em 25 de outubro de 2011.

HEYMANN, Luciana. *Um olhar antropológico sobre o documento: representações e usos sociais*. In: FREITAS, Lidia Silva de; MARCONDES, Carlos H.; RODRIGUES, Ana Célia (Orgs.). DOCUMENTO: gênese e contextos de uso. Niterói: PPGCI, EDUFF, 2010. ISBN 978-85-228-0638-6. p. 111-122.

HJORLAND, B. Discussion of Dahlberg's theory of concepts and knowledge organization. 2007. Disponível em: http://www.db.dk/bh/Lifeboat_KO/CONCEPTS/discussion_of_dahlberg.htm.

HJORLAND, Birger. Information seeking and subject representation: an activity-theoretical

approach to Information Science. New York: Greenwood Press, 1997.

HJORLAND, Birger. *Principia Informatica. Foundational Theory of Information and Principles of Information Services*. In: Emerging Frameworks and Methods. Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS4). Colorado, EUA. 2002. Disponível em: http://www.iva.dk/bh/core%20concepts%20in%20lis/articles%20a-z/principia_informatica.htm. Acesso em 10 de maio de 2010.

HJORLAND, Birger. Theory and metatheory of Information Science: a new interpretation. Journal of Documentation, v. 54, n. 5, dez. 1999.

HOBSBAWM, Eric. A era dos extremos: o breve século 20 (1914-1991). São Paulo: Companhia das Letras, 1995.

HORRIDGE, Matthew; MOULTON, Georgina; RECTOR, Alan; STEVENS, Robert; WROE, Chris. A practical guide to building OWL Ontologies using Protégé 4 and CO-ODE Tools. The University of Manchester, 2007.

HYVÖBEN, Eero; ALM, Olli; KUITINEN, Heini. Using an Ontology of Historical Events in Semantic Portals for Cultural Heritage. 2007. Disponível em <http://museosuomi.cs.helsinki.fi/publications/2007/hyvonen-et-al-history-2007.pdf>. Acesso em 25 de maio de 2010.

ICOM/CIDOC CRM Special Interest Group. Definition of the CIDOC Conceptual Reference Model. Version 5.0.4, November 2011. Disponível em: http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf.

IDE, N.; WOOLNER, D. "*Historical Ontologies*", in: Ahmad, K, Brewster, C., and Stevenson, M. (eds.), Words and Intelligence II: Essays in Honor of Yorick Wilks, Springer, 137-152, 2007. Disponível em: http://www.cs.vassar.edu/~ide/papers/festschrift.pdf

KANEIWA, Ken; IWAZUME, Michiaki; FUKUDA, Ken. *An Upper Ontology for Event Classifications and Relations*. In Proceedings of Australian Conference on Artificial

Intelligence. 2007. Disponível em http://www2.nict.go.jp/x/x163/kaneiwa/kk-ai07.pdf. Acesso em 23 de abril de 2011.

KANT, Immanuel. Crítica da razão pura. Trad. Alexandre Fradique Morujao. 5. Ed. Lisboa: Fundação Calouste Gulbenkian, 2001.

KASCHEK, Roland. *Issues of a Conceptual Modelling Agenda*. In: KASCHEK, Roland; DELCAMBRE, Louis (eds). The Evolution of Conceptual Modelling. Springer-Verlag Berlin Heidelberg, 2011.

KAULA, Prithivi. *Repensando os conceitos no estudo de classificação*. In: Conferência sobre pesquisa em classificação, 1982, Augsburg, ALE. Disponível em: http://www.conexaorio.com/biti/kaula/index.htm. Acessado em 3 de setembro de 2010.

KEET, Maria C. The *Use of Foundational Ontologies in Ontology Development: An Empirical Assessment*. In Proceedings of ESWC, 2011.

KOVACS, Katalin; DOLBEAR, Catherine; et al. *A methodology for building conceptual domain ontologies*. Technical report, Ordnance Survey Research and Innovation, 2006. Disponível em: http://www.ordnancesurvey.co.uk/partnerships/research/publications/docs/2006/Methodology_for_Building_Conceptual_Domain_Ontologies_V1.pdf. Acesso em 20 de dezembro de 2010.

LAKATOS, E. M.; MARCONI, M. A. Fundamentos de metodologia científica. 3. ed. São Paulo: Atlas, 1991.

LASSILA, Ora; MCGUINNESS, Debora. *The Role of Frame-Based Representation on the Semantic Web*. In Electronic Transactions on Artificial Intelligence, 2001. Disponível em: http://www-ksl.stanford.edu/people/dlm/etai/etai-abstract.html. Acesso em 20 de dezembro de 2010.

LE GOFF, Jacques. *Prefácio*. In: BLOCH, Marc. Apologia da história, ou, O ofício do historiador. Rio de Janeiro: Jorge Zahar Editor, 2001.

LEVY, Pierre. Cibercultura. São Paulo, Ed. 34. 1999.

LOEBE, F. *Abstract vs. social roles - a refined top-level ontological analysis*. In Procs. of AAAI Fall Symposium Roles'05, pages 93–100. AAAI Press, 2005. Disponível em: http://www.onto-med.de/publications/2007/loebe-f-2007-127-a.pdf

LUCAS, Clarinda Rodrigues. Leitura e interpretação em biblioteconomia. Campinas, SP, Ed. Unicamp. 2000.

MACEDO, D. C. *A construção do significado*. In: Saber Digital: Revista Eletrônica do CESVA, Valença, v. 1, n. 1, p. 70-84, mar./ago. 2008. Disponível em: http://www.faa.edu.br/revista/v1_n1_art04.pdf. Acesso em: 03 de janeiro de 2011.

MARCONDES, Carlos H. *Linguagem e documento: externalização, autonomia e permanência*. In: FREITAS, Lidia Silva de; MARCONDES, Carlos H.; RODRIGUES, Ana Célia (Orgs.). DOCUMENTO: gênese e contextos de uso. Niterói: PPGCI, EDUFF, 2010. ISBN 978-85-228-0638-6. p. 97-110.

MARCONDES, Carlos Henrique. *Um modelo semântico de publicações eletrônicas*. In: ENANCIB Encontro da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informaçao. Anais. Rio de Janeiro: PPGCI/IBICT, 2010.

MARCONDES, Carlos Henrique; CAMPOS, Maria Luiza de Almeida. *Ontologia e Web Semântica: o espaço da pesquisa em Ciência da Informação*. Pontodeacesso, Salvador, v. 2, n. 1, p.107-136, jun./jul. 2008. Disponível em: http://www.portalseer.ufba.br/index.php/revistaici/article/viewArticle/2669. Acesso em 01 de julho de 2010.

MARCONDES, Carlos Henrique; MENDONÇA, Marília Alvarenga Rocha; MALHEIROS, Luciana Reis; COSTA, Leonardo Cruz da; SANTOS, Tatiana. Cristina Paredes. *Ontologias como novas bases de conhecimento científico*. Perspectivas em Ciência da Informação, v13, n. 3. 2008. p.20-39.

MARTINS, A.F., FALBO, R.A; GUIZZARDI, G., ALMEIDA, J.P.A., *Uso de uma Ontologia de Fundamentação para Dirimir Ambiguidades na Modelagem de Processos de Negócio*. 7th

Brazilian Symposium on Information Systems, Salvador, Brazil, 2011. http://www.lbd.dcc.ufmg.br/colecoes/sbsi/2011/usodeumaantologia.pdf

MASOLO, C. et al. *Social roles and their descriptions*. Proceedings of the Ninth International Conference on the Principles of Knowledge Representation and Reasoning. AAAI Press, 2004.

MASOLO, C.; BORGO, S.; GANGEMI, A.; GUARINO, N.; OLTRAMARI, A; SCHNEIDER, L. '*WonderWeb Deliverable D17*. The WonderWeb Library of Foundational Ontologies and the DOLCE ontology' , Technical report, ISTC-CNR, 2002. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.11.4243

MASOLO, Claudio; BORGO, Stefano; GANGEMI, Aldo; GUARINO, Nicola; OLTRAMARI, Alessandro. Ontology Library. Laboratory For Applied Ontology – ISTC-CNR. Trento, 2003. Disponível em: http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf.

MASTELA, L. S.; ABEL, M.; LAMB L. C.; DE ROS, L. F. *Uma Ontologia Temporal para Modelagem de Conhecimento sobre Ordenação de Eventos*. XXV Congresso da Sociedade Brasileira de Computação. UNISINOS - São Leopoldo - RS. 2005. Disponível em: http://www.sbc.org.br/bibliotecadigital/download.php?paper=321 Acesso em 23 de agosto de 2010.

MERRILL, Gary H. *Ontological realism: Methodology or misdirection*? Applied Ontology, 5 (2010) 79–108. Disponível em: http://iospress.metapress.com/content/j3324564p5l33863/

MILLER, Eric. The semantic Web. 2004. Disponível em: http://www.w3.org/2004/Talks/0120-semweb-umich/Overview.html . Acesso em 16/05/2010.

MILLER, Eric; SWICK, Ralph. *An Overview of W3C Semantic Web Activity*. Bulletin of The American Society for Information Science and Technology, v. 29, n. 4, Apr/May 2003, pp 8-11.

MOREIRA, A.; ALVARENGA, L.; OLIVEIRA, A.P. *O nível do conhecimento e os instrumentos de representação: tesauros e ontologias*. Datagramazero, v.5, n.6, dez. 2004.

Disponível em: http://www.datagramazero.org.br/dez04/Art_01.htm. Acesso em 10 de maio de 2010.

MORRIS, Charles. Fundamentos da teoria dos signos. Rio de Janeiro: Ed. Eldorado; São Paulo: Ed. USP, 1976.

MOURA, Maria Aparecida. *Leitor-bibliotecário: interpretação, memória e as contradições da subjetividade em processos de representação informacional*. In: NAVES, M. e KURAMOTO, H. Organização da Informação: princípios e tendências. Brasília, DF: Brisquet de Lemos, 2006.

MURGUIA, Eduardo Ismael. *Documento e instituição: produção, diversidade e verdade*. In: FREITAS, Lidia Silva de; MARCONDES, Carlos H.; RODRIGUES, Ana Célia (Orgs.). DOCUMENTO: gênese e contextos de uso. Niterói: PPGCI, EDUFF, 2010. ISBN 978-85-228-0638-6. p. 123-140.

MYLOPOULOS, John. *Conceptual Modelling and Telos*. In: LOUCOPOULOS, P., ZICARI, R (Eds). Conceptual Modelling, Databases and Case: An Integrated View of Information System Development. Wiley, 1992. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.83.3647. Acesso em 14/11/2010.

NAGYPÁL, Gábor. *Creating an application-level ontology for the complex domain of history: Mission impossible*? In Proceedings of Lernen - Wissensentdeckung - Adaptivität (LWA 2004), FGWM 2004 Workshop, pages 287–294, Berlin, Germany, Oct. 4–6 2004. Disponível em: <http://www.fzi.de/images/files/pub/fgwm2004.pdf >. Acesso em: 3 de janeiro de 2011.

NAGYPÁL, Gábor. *History ontology building: the technical view*. In: INTERNATIONAL CONFERENCE OF THE ASSOCIATION OF HISTORY AND COMPUTING, 16, 2005, Proceedings... Amsterdam: Royal Netherlands Academy of Arts and Sciences, 2005a. Disponível em <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.103.2193>. Acesso em 30 de maio de 2010.

NAGYPÁL, Gábor; DESWARTE, Richard; OOSTHOEK, Jan. *Applying the Semantic Web: The VICODI Experience in Creating Visual Contextualization for History*. Literary and Linguistic Computing, v. 20, n. 3, 2005. Disponível em: <http://semweb.weblog.ub.rug.nl/sites/semweb.weblog.ub.rug.nl/files/327_0.pdf>. Acesso em 29 de maio de 2010.

NAGYPÁL, Gábor; MOTIK, B. *A fuzzy model for representing uncertain, subjective, and vague temporal knowledge in ontologies*. In R. Meersman, Z. Tari, and D. C. Schmidt, editors, On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE, volume 2888 / 2003 of Lecture Notes in Computer Science, pages 906 – 923. Springer-Verlag. 2003. Disponível em: <http://www.vicodi.org/odbase2003.pdf>. Acesso em: 3 de janeiro de 2011.

NEDOBITY, W. Concepts versus meaning as reflected by the works of Wuster and Wittgenstein. International Classification. V. 16, n. 1, p. 24-26, 1989.

NOY, Natalya F.; McGUINNESS, Deborah L. Ontology development 101: a guide to create your first ontology. Stanford University, Stanford, CA. 2001. Disponível em: http://protege.stanford.edu/publications/ontology_development/ontology101.html. Acesso em: 27 de novembro de 2010.

OBERLE, Daniel. et al. *Dolce ergo sumo: On foundational and domain models in swinto (smartweb integrated ontology).* In Proceedings of 3rd International Conference on Language Resources and Evaluation Conference (LREC) and OntoLex workshop. 2006. Disponível em: http://people.aifb.kit.edu/phi/pub/SWIntO.pdf. Acesso em 22/10/2011.

OMG OBJECT MANAGEMENT GROUP. OMG Unified Modeling Language (OMG UML): Infrastructure Specification. 2011. Disponível em: http://www.omg.org/spec/UML/2.4.1/Infrastructure/PDF.

PATTUELLI, Maria Cristina. A user-centered approach to the development of a history domain ontology: helping teachers use digital primary sources. University of North Carolina. Dissertation, 2007. Disponível em: http://gradworks.umi.com/32/72/3272756.html.

NILES, Ian; PEASE, Adam. *Towards a Standard Upper Ontology*. In: Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001). Maine, 2001. Disponível em: http://pdf.aminer.org/000/212/663/towards_a_standard_upper_ontology.pdf

PROBST, Florian. *Ontological Analysis of Observations and Measurements*. In Proceedings of the 4th International Conference on Geographic Information Science (GIScience). Münster, Alemanha. 2006.

RADEMAKER, Alexandre. A Proof Theory for Description Logics. London: Springer, 2012. http://dx.doi.org/10.1007/978-1-4471-4002-3.

RADEMAKER, Alexandre; PAIVA, Valeria de. *Revisiting a Brazilian WordNet*. In: Global Wornet Conference 2012, 2012, Matsue. Proceedings of Global Wordnet Conference. Matsue: Global Wordnet Association, v. 1. p. 100-104, 2012. Disponível em: http://www.globalwordnet.org/gwa/proceedings/gwc2012.pdf. Acesso em 17 de junho de 2012.

RAGHAVAN, Prabhakar. Finding anything in the billion page Web: are Algorithms the key? (Panel Abstract), Toronto, Canada, Elsevier Science, 1999. Disponível em http://www.ra.ethz.ch/CDstore/www8/data/2196/pdf/pd1.pdf. Acesso em 27 de março de 2011.

RAIMOND, Yves; ABDALLAH, Samer. The Event Ontology. 2007. Disponível em: http://motools.sourceforge.net/event/event.html

RANDALL, Davis; SHROBE, Howard; SZOLOVITS, Peter. *What is a Knowledge Representation*? AI Magazine, 14(1): 17-33, 1993. Disponível em http://www.aaai.org/ojs/index.php/aimagazine/article/view/1029/947. Acesso em 14 de novembro de 2011.

RIBEIRO, Fernanda. *Processo natural ou arranjo a posteriori*? Leituras: Revista da Biblioteca Nacional. Série 3. V.2, Lisboa, out. 1997-abr. 1998.

RICOEUR, P. Time and Narrative. Volume 1. Chicago and London, 1984.

ROBERTSON, Bruce. Introduction to HEML. The Historical Event Markup and Linking Project. 2005. Disponível em: http://heml.mta.ca/heml-cocoon/description. Acesso em 18 de maio de 2011.

ROBERTSON, Bruce. *Visualizing an historical semantic web with HEML*. In Proceedings of the 15th International Conference on World Wide Web (Edinburgh, Scotland, May 23 - 26, 2006). WWW '06. ACM Press, New York, NY.

ROBIN, Régine. História e Linguística. São Paulo, Cultrix. 1973.

SALES, L. F. ; CAMPOS, M. L. A. ; GOMES, H. E. *Ontologias de domínio: um estudo das relações conceituais*. Perspectivas em Ciência da Informação, v. 13, p. 62-76, 2008.

SARACEVIC, T. *Ciência da informação: origem, evolução e relações*. Perspectivas em Ciência da Informação, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

SAYÃO, L. F. *Modelos teóricos em ciência da informação: abstração e método científico*. Revista Ciência da Informação, Brasília, v. 30, n. 1, jan./abr. 2001. Disponível em: http://revista.ibict.br/index.php/ciinf/article/viewArticle/228. Acesso em 14/11/2010.

SCHERP, A; FRANZ, T; SAATHOFF, C; STAAB, S. A model of events based on a foundational ontology. Alemanha: Fachbereich Informatik, n. 2, 2009. Disponível em: http://kola.opus.hbz-nrw.de/volltexte/2009/359/pdf/2009_02_Arbeitsberichte.pdf

SCHERP, Ansgar; STAAB, Steffen; FRANZ, Thomas; SAATHOFF, Carsten. A model of events for integrating event-based information in complex socio-technical information systems. Slides presentation. Institute WeST, University of Koblenz, 2012. Disponível em: http://www.slideshare.net/ascherp/a-model-of-events.

SEWELL JR., Willian H. Historical events as transformations of structures: inventing revolution at the Bastille. Theory and Society, v. 25, n. 6, p. 841-881, Dec. 1996. Disponível em http://ic.ucsc.edu/~traugott/hist171/readings/Sewell%20Bastille.pdf. Acesso em 12/06/2010.

SHAW, Ryan; TRONCY, Raphael; HARDMAN, Lynda. LODE: Linking Open Descriptions of Events. Technical report. UC Berkeley, 2009. Disponível em http://oai.cwi.nl/oai/asset/14783/14783A.pdf

SCHWARCS, Lilia Moritz. *Por uma historiografia da reflexão (Prefácio).* In: BLOCH, Marc. Apologia da história, ou, O ofício do historiador. Rio de Janeiro: Jorge Zahar Editor, 2001.

SILVA, Daniela Lucas da. Uma Proposta Metodológica para Construção de Ontologias: Uma Perspectiva Interdisciplinar entre as Ciências da Informação e da Computação. Dissertação. UFMG, Belo Horizonte. 2008.

SILVA, Daniela Lucas da; SOUZA, Renato Rocha; ALMEIDA, Maurício Barcellos. *Ontologias e vocabulários controlados: comparação de metodologias para construção*. Ci. Inf., Brasília, v. 37, n. 3, 2008. Disponível em http://www.scielo.br/scielo.php?pid=S0100-19652008000300005&script=sci_arttext. Acesso em 07/06/2010.

SILVA, Stanley Plácido da Rosa. *O ofício do historiador e a escrita da História: dos Annales à pós-modernidade, da História às "histórias"*. In: Revista Litteris. 2009. Disponível em: http://revistaliter.dominiotemporario.com/doc/oficiodohistoriadoreaescr.pdf. Acesso em 24/07/2012.

SMITH, Barry. *Beyond Concepts, or: Ontology as Reality Representation*. Achille Varzi and Laure Vieu (eds.), Formal Ontology and Information Systems. Proceedings of the Third International Conference (FOIS 2004), Amsterdam: IOS Press, 2004.

SMITH, Barry. Ontology and Information Systems [online]. 2002. Disponível em: http://ontology.buffalo.edu/ontology(PIC).pdf . Acesso em 29/05/2010.

SMITH, Barry. *Ontology as Reality Representation*. Achille Varzi and Laure Vieu (eds.), Formal Ontology and Information Systems. Proceedings of the Third International Conference (FOIS 2004), Amsterdam: IOS Press, 2004. Disponível em: http://ontology.buffalo.edu/bfo/Beyond_Concepts.pdf

SMITH, Barry; CEUSTER, Werner. *Ontological realism: A methodology for coordinated evolution of scientific ontologies*. Applied Ontology, 5, 2010. Disponível em: http://iospress.metapress.com/content/1551884412214u67

SMITH, Carol. *Historians and information: Patterns of collection, cognition and need. A Review of the Literature.* Info 511. 2004a. Disponível em: http://www.carolsmith.us/downloads/511rol.pdf

SMITH, Michael K.; WELTY, Chris; McGUINESS, Deborah L. OWL Web Ontology Language Guide. 2004. Disponível em: http://www.w3.org/TR/owl-guide/. Acesso em 10 de dezembro de 2010.

SOUZA, Renato Rocha, CAIXETA, Mario; *Representação do conhecimento: história, sentimento e percepção*. Informação & Informação. v. 13, n. 2, 2008. Disponível em: http://www.brapci.ufpr.br/download.php?dd0=11367. Acesso em 10 de novembro de 2010.

SOUZA, Renato Rocha. *Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências*. Perspect. ciênc. inf., Belo Horizonte, v. 11, n. 2, 2006 . Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362006000200002&lng=en&nrm=iso. Acesso em 27 de Junho de 2010.

SOUZA, Renato Rocha; ALVARENGA, Lídia. *A Web Semântica e suas contribuições para a ciência da informação*. Ci. Inf., Brasília, v. 33, n. 1, Apr. 2004 . Disponível em http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652004000100016&lng=en&nrm=iso Acesso em 12/05/2010.

SOUZA, Renato Rocha; HIGUCHI, Suemi; RADEMAKER, Alexandre; SÁ, Asla Medeiros. *Portal Semântico do CPDOC*. In: Anais do IV Encontro de Bases de Dados Sobre Informações Arquivísticas, Rio de Janeiro, 2011.

SOWA, John F. Building, sharing and merging ontologies. Tutorial. 1999. Disponível em: http://www.jfsowa.com/ontology/ontoshar.htm. Acesso em 23 de outubro de 2010.

SOWA, John F. Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole Publishing Co., Pacific Grove, CA, 2000.

STUDER, Rudi; BENJAMINS, Richard; FENSEL, Dieter. *Knowledge engineering: principles and methods.* Data and Knowledge Engineering 25, 1998. Disponível em: http://www.it.iitb.ac.in/~palwencha/ES/Knowledge%20engineering%20-%20Principles%20and%20methods.pdf

TALJA, S. *Constituting 'information' and 'user' as research objects: a theory of knowledge formations as an alternative to the information-man theory*. In P. Vakkari, R. Savolainen & B. Dervin (eds). Information Seeking in Context. London: Taylor Graham: 67-80. 1997. http://www.uta.fi/~lisaka/ISIC.HTM

TAMBA-MECZ, Irene. A Semântica. Trad. Marcos Marcionilo. São Paulo: Parábola Editorial, 2006.

USCHOLD, Mike. Ontologies: principles, methods, and applications. Knowledge Engineering Review 11, 1996.

VAN HAGE, W; MALAISÉ, V; de VRIES, G; SCHREIBER, G; VAN SOMEREN, M. *Combining ship trajectories and semantics with the Simple Event Model (SEM).* In: 1st ACM International Workshop on Events in Multimedia (EiMM 2009). Beijing, China, 2009. Disponível em: http://www.few.vu.nl/~WR.van.Hage/papers/eimm30070-vanhage.pdf

VAN HAGE, Willem; MALAISÉ, Véronique; SEGERS, Roxane; HOLLINK, Laura; SCHREIBER, Guus. *Design and use of the Simple Event Model (SEM).* Journal of Web Semantics, Elsevier, 2011. http://www.cs.vu.nl/~guus/papers/Hage11b.pdf

VELTMAN, Kim. Cultural and Historical Metadata: MEMECS (Metadonnées et Mémoire Collective Systématique) . July 2000. Disponível em http://www.cultivate-int.org/issue1/memecs/. Acesso em 25/10/2010.

VICKERY, B.C. Faceted Classification - a guide to construction and use of special schemes. London: Aslib, 1960.

VICKERY, D.C. Knowledge representation: a brief review. Journal of documentation, v.42, n.3, sept. 1986.

W3C. DUL ssn documentation. [Versão de] 31 de maio de 2011. Disponível em: http://www.w3.org/2005/Incubator/ssn/wiki/DUL_ssn#Event

W3C. Time Ontology in OWL. Working Paper. 2006. Disponível em: http://www.w3.org/TR/owl-time/. Acessado em 21 de fevereiro de 2011.

WEINSTEN, Peter. *Ontology-based metadata: transforming the MARC legacy*. The Third International ACM Conference on Digital Libraries, 1998. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.44.5722. Acesso em 22 de fevereiro de 2011.

WIERINGA, Roel. *Real-World Semantics of Conceptual Models*. In: KASCHEK, Roland; DELCAMBRE, Louis (eds). The Evolution of Conceptual Modelling. Springer-Verlag Berlin Heidelberg, 2011.

WÜESTER, E. *L'étude scientifique générale de la terminologie, zone frontalière entre la Linguistique, la Logique, l'Ontologie, l'Informatique et les Sciences des Choses*. In: RONDEAU, G.; FELBER, F. (Org.). Textes choisis de terminologie: fondéments théoriques de la terminologie. Québec: GIRSTERM, 1981. p. 57-114.

ZHOU, Q.; FIKES R. *A reusable time ontology*. Technical report, Knowledge. Stanford Laboratory, 2001. Disponível em: http://www.aaai.org/Papers/Workshops/2002/WS-02-11/WS02-11-015.pdf. Acessado em 21 de fevereiro de 2011.